Summer 2017

Development and Test of a Theory of Applicant Reactions to Branched Situational Judgment Tests

Craig Matthew Reddock
Old Dominion University, credd005@odu.edu

DEVELOPMENT AND TEST OF A THEORY OF APPLICANT REACTIONS TO BRANCHED SITUATIONAL JUDGMENT TESTS

by

Craig Matthew Reddock
B.A. May 2003, Oglethorpe University
M.S. May 2009, University of Tennessee Chattanooga

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

INDUSTRIAL/ORGANIZATIONAL PSYCHOLOGY

OLD DOMINION UNIVERSITY
August, 2017

Approved by:

Richard N. Landers (Director)

Debra A. Major (Member)

Ryan L. Klinger (Member)

ABSTRACT

DEVELOPMENT AND TEST OF A THEORY OF APPLICANT REACTIONS TO BRANCHED SITUATIONAL JUDGMENT TESTS

Craig Matthew Reddock
Old Dominion University, 2017
Director: Dr. Richard N. Landers

Branched situational judgment tests (BSJTs) are a recent innovation in personnel selection. Yet, there is little research supporting their use. Data from a qualitative pilot study indicated that BSJTs are expected to result in more positive applicant reactions. Using a grounded theory approach, Study 1 developed a theoretical model of the impact of branching on applicant reactions. Several perceived procedure characteristics were theorized to mediate the relationship between both the general use of branching as well as different branching features and applicant perceptions. Study 2 empirically tested the model developed in Study 1. Results indicated that consistency of administration mediated the relationship between general branching and motivation and the relationship between general branching and procedural justice. Although branching in general resulted in increased motivation, it resulted in decreased perceptions of fairness. Consistency of administration also mediated the relationship between the use of parallel branching, a specific branching feature, and procedural justice, as well as between parallel branching and test attitude. BSJTs that utilized parallel branches resulted in more positive perceptions of fairness and test attitude. Looping had a positive impact on perceptions of fairness, though this relationship was not mediated by opportunity to perform. Limitations and practical implications are discussed.

This dissertation is dedicated to my wife, Crysta, and daughter, Amelia

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I
INTRODUCTION

Recent work on situational judgment tests (SJTs) has seen the emergence of a new kind of assessment: the branching SJT (BSJT). This new type of assessment first appeared in the scholarly literature when Olson-Buchanan, Drasgow, Moberg, Mead, Keenan, and Donovan (1998) described the creation and validation of a conflict resolution assessment in which participant responses to questions on a video SJT led different participants to different videos. Since the publication of that study, it appears that the use of BSJTs for developmental and personnel selection purposes has increased. Several large human resources consulting companies now offer BSJTs as part of their commercial products (e.g., HumRRO, 2014). These assessments are delivered in a variety of forms, including static text, interactive video, and animation (e.g., Chan & Schmitt, 1997, Lievens & Sackett, 2006; Richman-Hirsch, Olson-Buchanan, & Drasgow, 2000). This increase in the use of BSJTs among practitioners is due to several perceived benefits, including increased engagement, increased face-validity, and increased test security. Practitioners report that clients are increasingly requesting assessments like BSJTs because test-takers find them more enjoyable and have more positive reactions to them than other assessment types. Branching more closely mirrors how real-world situations might unfold, given the test-takers' responses. Specifically, once a test-taker chooses a response, the next scenario provided to that test-taker is informed by that response, resulting in a more coherent, more story-like assessment. Finally, because the assessment content varies somewhat among test-takers, test security is likely to be enhanced.

Despite the popularity of BSJTs, there are several theoretical issues surrounding BSJTs that require attention. Very little existing research literature has explored why practitioners adopt BSJTs. Understanding more about why practitioners opt to use BSJTs is critically important in fully understanding how these assessments are constructed, how they are psychometrically evaluated, and for what purposes they may be best suited. Furthermore, there is little published information regarding how BSJTs may branch, what triggers branching, and how branching may impact measurement. This is not to say that there is not a thoughtful, scientifically sound process underpinning the creation of BSJTs, but

rather that such information is not widely shared. Given the increased use of BSJTs in applied settings, tapping the knowledge base of practitioners regarding BSJTs is important to building our scientific knowledge of such tests.

To justify the continued use of BSJTs, the new theory suggested by these practitioner assumptions and related literature must also be empirically tested. To date, there appears to be only three peer-reviewed journal articles on BSJTs (Kanning, Grewe, Hollenberg, & Hadouch, 2006; Olson-Buchanan et al., 1998; Richman-Hirsch et al., 2000). While informative, much more research is needed to understand how important psychometric and psychological concepts (e.g., validity, reliability, applicant reactions) are impacted by the use of BSJTs. Related research literature provides some guidance here; for example, research on computer adaptive testing (CAT) suggests that the way adaptive tests are constructed can impact perceived-performance and feedback acceptance (Tonidandel, Quinones, & Adams, 2002). Roch, Mishra, and Trombini (2014) found a relationship between semi-structured interviews and enhanced applicant motivation, whereas Schleicher, Venkataramani, Morgeson, and Campion (2006) found that using semi-structured interviews was likely to give applicants a greater sense of having the opportunity to perform. It is critical to determine if BSJTs, given the variability in their administration, are viewed similarly.

The purpose of the present study is, therefore, to fill this gap in the literature by developing this needed theory and empirically exploring BSJTs. A qualitative pilot study was conducted to inform the most fruitful directions for this effort. Study 1 qualitatively examined BSJT deployment in the field in order to develop a theory of SJT branching. Study 2 quantitatively examined the effects upon applicant reactions suggested by this theory. Thus, this is mixed method research combining both quantitative and qualitative elements, which is ideal because it allows for deeper elaboration of phenomena, increases generalizability, collects triangulating evidence, and facilitates complex interpretation (Gibson, 2016).

Situational Judgment Tests

SJTs are a measurement method that presents applicants with a scenario and asks them how they would respond to that scenario (Ployhart & Ward, 2013). SJTs typically involve presentation of a variety of situations applicants might encounter on the job, with each situation followed by several possible courses of action. The test-taker then indicates which of these options they would or should do (Weekley & Ployhart, 2006). Intended to measure judgment in work settings (McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001), SJTs measure a combination of work-related knowledge, skills, and abilities (Weekley, Hawkes, Guenole, & Ployhart, 2015).

McDaniel et al. (2001) briefly described the history of SJTs, which dates back as far as the 1920s, when the George Washington Social Intelligence Test included subtests called Judgment in Social Situations. During World War II, army psychologists used tests to assess the judgment of soldiers, whereas in the 1940s, tests intended to assess supervisory potential began to proliferate (McDaniel et al., 2001). Use of SJTs in various forms continued throughout the following decades, but it was not until the publication of Motowidlo, Dunnette, and Carter's (1990), article on the low fidelity simulations that research on SJTs increased "geometrically" (Ployhart & MacKenzie, 2011), resulting in a vibrant research stream that continues today.

SJTs have several strengths as predictors of job performance that make them attractive to practitioners. SJTs have been used to assess a variety of constructs, including personal initiative (Bledlow & Frese, 2009), work habits (Chan & Schmitt, 1997), and interpersonal skills (Chan & Schmitt; Lievens & Sackett, 2006; Lievens, Buyse, & Sackett, 2005; Lievens, Sackett, & Buyse, 2009), to name just a few. Additionally, they have shown low to moderate meta-analytic correlations with cognitive ability ($\rho = .32$) and all of the Big Five facets of extroversion ($\rho = .14$), agreeableness ($\rho = .25$), conscientiousness ($\rho = .27$), emotional stability ($\rho = .22$), and openness to experience ($\rho = .13$; McDaniel, Hartman, Whetzel, & Grubb, 2007). SJTs as a method demonstrate moderate validity, (mean $r = .20$; McDaniel et al., 2007), while manifesting smaller subgroup differences compared to cognitive ability tests (Clevenger,

Pereira, Wiechmann, Schmitt, & Harvey, 2001). Additionally, SJTs demonstrate incremental validity over cognitive ability, personality, and experience in the prediction of job performance (Clevenger et al., 2001; Weekley & Ployhart, 2005).

SJTs have some significant weaknesses as well, which often limit their practicality. It was not until relatively recently that researchers understood more about what was being measured by SJTs and why they work (Lievens & Motowidlo, 2016). While SJTs demonstrate smaller subgroup differences than cognitive ability tests, they have larger subgroup differences than non-cognitive predictors such as personality measures and interviews (Ployhart & Holtz, 2008). Because job experts are typically needed to help craft realistic scenarios, SJTs can be somewhat costly to create (Ployhart & MacKenzie, 2011), particularly if interactive media, such as video and game-like interfaces, are used. Relatedly, because creating alternate forms is often costly, test security is a concern (Ployhart & MacKenzie, 2011).

Branching Situational Judgment Tests

As noted earlier, a relatively recent innovation in the SJT research domain was the introduction of branching. BSJTs work the same way as traditional SJTs – a work-related scenario is presented to the test-taker, along with several response options. However, unlike SJTs, the scenario sequence is not the same for all test-takers. The scenarios that each test-taker sees depends, at least in part, on the test-takers' responses. Consider the following scenario:

Recently one of your employees has had difficulty turning in assignments on time. In addition to this, the work the employee is turning in contains a large number of easily preventable errors. Previously, this individual was a model employee who always met deadlines and turned in high quality work. This employee can occasionally be sensitive to criticism. What would you do?

a. Take the employee aside for an informal conversation and ask them if there is anything that is bothering them.
b. Contact Human Resources and ask them for guidance on how to proceed.
c. Wait and see if the problem persists, and if it does, then decide how to proceed.

d. Request a formal meeting with the employee and discuss the recent performance issues.

Note that the response options detailed above all have consequences for how the situation might unfold in a real life situation. With a BSJT, one or more of the responses could lead to a completely different scenario than the others. That is, if someone chose response *b*, the next scenario they see might involve a conversation with Human Resources. However, if a test-taker chose response *d*, the next scenario might involve a conversation with the employee.

In one of the first empirical studies of BSJTs, Olson-Buchanan et al. (1998) developed and validated a BSJT measuring conflict resolution skills. An interactive video was created that contained nine conflict scenarios, each with 4 response options, with each option branched to a different scenario. Every scenario, option, and branch was created to be realistic and in accordance with the conflict resolution model used by the authors. This was achieved by collecting scenarios through structured interviews, collecting possible responses to these scenarios through surveys sent out to managers, collecting ratings of those response options, and linking the resulting scenarios and response options to the conflict resolution model. After the nine main scenarios were developed, a branch for each response option was created, which the authors noted increased the interactivity and the realism of the assessment. All of the scenes were scripted and videotaped for use in the study.

Olson-Buchanan et al.'s (1998) study design enabled generalization across a wide group of potential test-takers. They collected six samples (N = 347) for the validation study, including managers and supervisors from a large Midwestern telecommunications company (N = 75), managers and supervisors from a variety of companies and industries (N = 136), members of a business manager's group from a large Midwestern university (N = 44), first line supervisors from a private sector printing company (N = 26), administrators, supervisors, and managers from a large Midwestern university (N = 23), and hospital managers (N = 43). Subjects completed computer adaptive assessments of verbal and quantitative skills, as well as the conflict resolution assessment. Job performance was measured via a composite of supervisor ratings of conflict resolution and performance.

Results showed that the conflict resolution measure was significantly related to manager ratings of on-the-job conflict resolution across several different scoring methods (cross validated r = .26 for both a hybrid and empirical keyed assessment), as well as to ratings of overall performance (r = .20 for the empirically keyed assessment, r = .13 for the hybrid keyed assessment). The conflict resolution BSJT was not significantly correlated with either the test of verbal skills or the test of quantitative skills. The conflict resolution assessment demonstrated significant incremental validity over the skills tests for both the conflict resolution criteria (r increased from .08 to .27) and overall ratings of job performance (r increased from .16 to .25).

Richman-Hirsh et al. (2000) used the same assessment developed by Olson-Buchannan et al. (1998) to examine differences in applicant reactions to three different assessment media: paper and pencil, a text-based version administered via a computer, and a multi-media version administered via a computer. The authors found that the multimedia version was rated as more content valid, more predictively valid, as having more job-relevant information, more enjoyable, shorter, and test-takers were more satisfied with the assessment process compared to the other two conditions.

In a more recent article, Kanning et al. (2006) described two studies they conducted comparing reactions between traditional and branching SJTs. Study 1 was a 2 (BSJT vs SJT responses) x 2 (video vs text-based responses) design, with typical scenarios that a police officer would encounter on the job (e.g., dealing with rowdy sports fans, dealing with a dispute between two people) as the stimuli. In the SJT, text-based condition, subjects watched a video then had to choose between two possible text-based responses. The BSJT condition was identical, except that the answer chosen by the respondent determined the content of a second video clip, again with two possible responses. Policemen rated the scenarios in terms of four criteria: usefulness, transparency, emotional reaction, and job-relatedness. The authors found that the BSJT items were rated as significantly higher than the SJT items only in terms of usefulness. There were no differences in ratings between video and non-video responses. The final hypothesis of Study 1, that a BSJT with video-based responses would be rated more positively in terms of

usefulness, emotional reaction, transparency, and job relatedness than the other three combinations, was not supported for any of the four outcomes.

Study 2 was the same as Study 1, except that Study 2 was a within-subjects design where the subjects saw all of the different item formats. Additionally, along with video-based responses, video- and text-based stimuli were introduced. Eighty-two policemen then rated each scenario in terms of five criteria: usefulness, transparency, emotional reaction, job-relatedness, and fairness. Additionally, subjects were asked to rank the six different conditions along the five criteria. That is, the six scenarios were ranked five times, once for each criterion. Results showed that the BSJT with video-based stimuli and video-based responses was rated significantly higher than the other scenarios in terms of usefulness, acceptance, and job-relatedness. For the fairness criterion, the BSJT with video-based stimuli and video-based responses was not rated significantly higher than the SJT with video-based stimuli and text-based responses and the SJT with video-based stimuli and video-based responses. In terms of ranking, the BSJT with video-based stimuli and video-based responses was ranked the highest across all 5 criteria. Taken together, these results indicated that BSJT items that use video for both the stimulus and response options elicit more favorable reactions than do other item types. While this study utilized a rather simple BSJT, it provides at least preliminary evidence that BSJTs have a positive impact on applicant reactions.

These three studies, which encompass almost all published BSJT research to date, all considered branching as a single characteristic of an SJT: branched or not branched. However, BSJTs are much more complex, and the implications of the various types of BSJTs are still relatively unknown. Olson-Buchanan et al. (1998), Richman-Hirsch et al. (2000), and Kanning et al. (2006) describe a simple type of branching, but this type is not representative of all approaches to branching. In them, a subject chooses a response, and based on that response, different scenarios appear. There are many more ways that SJTs can branch. Weekley et al. (2015) describe two ways that BSJTs can branch: nonlinear and pseudo-non-linear. In a nonlinear BSJT, each response option leads to a different scenario. As the authors point out, in such a BSJT, the number of scenarios would increase exponentially, leading to some practical issues (e.g., cost,

test size). In a pseudo-linear BSJT, response options lead to logical cut-scenes, but then branching converges and all participants see the same follow-up item. These may be just a few potential examples of many. Thus, one goal of this research is to attempt to discover the different ways that SJTs can branch.

Research Question 1 (RQ1). What are the ways that SJTs can branch?

Given that SJTs can branch in many different ways, it is important to understand the particular mechanisms and approaches underlying how BSJT creators believe branching impacts applicant reactions, as well as their beliefs regarding how different kinds of branching impacts applicant reactions. For example, some practitioners may implement branching in a particular way for particular reasons whereas other practitioners may implement branching in different ways for different reasons. These practitioners may furthermore measure reactions to their assessments differently given their differing BSJT design approaches. Thus, a major goal here is to understand these perspectives more completely in order to create a more unified theory of branching.

Research Question 2 (RQ2). What are the different ways that branching impacts applicant reactions?

Research Question 3 (RQ3). How is branching used to impact applicant reactions?

Research Question 4 (RQ4). How are applicant reactions uniquely measured for branching?

Applicant Reactions Theory

An important goal of those practicing personnel selection is maximizing applicant reactions to hiring measures. While the primary goal of any selection system is to make accurate hiring decisions, ensuring that the selection process does not have a negative impact on a job candidate's perception of the hiring organization or their performance on the selection measures is also of critical importance (Ryan & Huth, 2008). Hausknecht, Day, and Thomas (2004) provide several specific reasons that applicant reactions are important. First, applicants who have a negative experience during the selection process may view the hiring organization negatively, and may also be less likely to accept a job offer. Second, candidates with negative experiences may tell others about their experience, potentially dissuading

potential applications. Third and finally, candidates who have negative experiences during the selection process may be more inclined to bring legal action against the offending organization. Truxillo, Bauer, and McCarthy (2015) categorized these outcomes as either soft (e.g., organizational attractiveness, recommendation intentions) or hard (e.g., litigation). Chan (1997) further found that applicant reactions have an impact on actual personality and cognitive ability test scores, while Chan and Schmitt (1997) found the same for SJT scores. Schmitt and Ryan (1997) found that test-taking attitudes were predictive of applicant withdrawal from the selection process, while Macan, Avedon, Paese, and Smith (1994) found that applicant reactions to selection tests positively predicted satisfaction with the process, the job, the organization, and acceptance intentions. Bauer, Maertz, Dolen, and Campion (1998), in a longitudinal study of applicant reactions, found that perceptions of justice impact applicant reactions to organizational attractiveness, intentions toward the company, and perceptions of testing fairness. Thus, applicant reactions have an impact on a wide range of important organizational outcomes (Hausknecht et al., 2004).

Hausknecht et al. (2004) developed one of the most comprehensive models of applicant reactions. Developed from an adaptation of Gilliland's (1993) model and Ryan and Ployhart's (2000) model with a meta-analysis of 86 independent samples (N = 48,750), the Hausknecht et al. model details how applicant perceptions of the selection process can predict important organizational outcomes through several key processes. Specifically, the authors hypothesized that person characteristics, perceived procedure characteristics, job characteristics, and organizational context all directly affect application perceptions, which in turn impact various outcomes. Perceived procedure characteristics also directly affect outcomes.

The present study focused on two specific part of two specific parts of Hausknecht et al.'s (2004) model: antecedents (person characteristics and perceived procedure characteristics) and applicant perceptions. Applicant perceptions are the key dependent variables, which include anxiety, perceptions of justice, and motivation. These perceptions are impacted by classes of antecedents, including: test-taker characteristics such as demographics and personality; perceived procedure characteristics such as justice rules, procedure length, and transparency; job characteristics such as knowledge, skills, and ability needed

for the job, industry norms for selection; and organizational context. Applicant perceptions in turn impact outcomes such as organizational attractiveness and perceptions of fairness. Figure 1 depicts these portions of the Hausknecht et al. model in a path diagram. As in Hausknecht et al.'s model, person characteristics impact applicant perceptions through perceived procedural characteristics, but person characteristics also have a direct impact on applicant perceptions. Based on the extant literature, these parts of the model were most likely to impact, or be impacted by, branching. That is, the most proximal impact of branching would be on perceived procedure characteristics (e.g., consistency of administration, face validity) and the resultant change in applicant perceptions, potentially moderated by person characteristics. The initial development of this model is explained in greater detail in the next section.

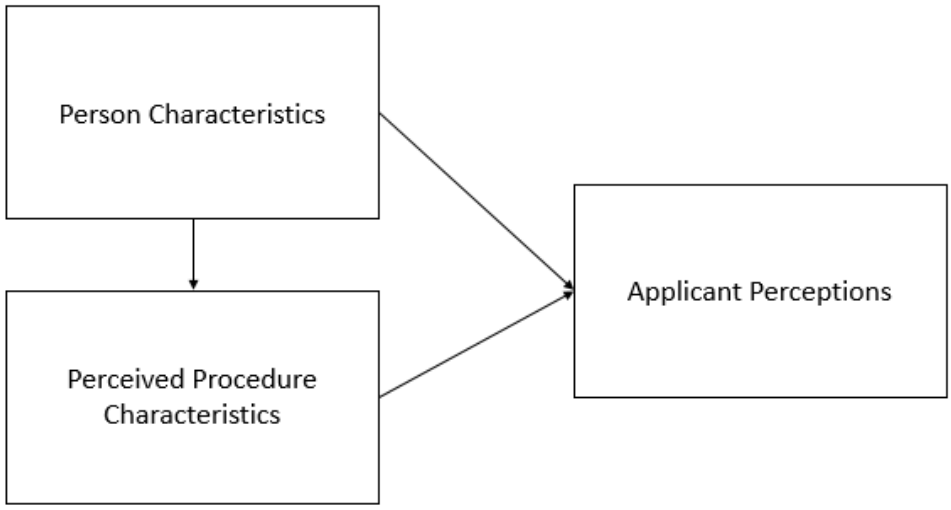

Figure 1. Subset of Hausknecht et al.'s (2004) model relevant to the present dissertation.

Applicant Reactions to Branched Situational Judgment Tests

Applicant perceptions are impacted by two types of antecedents. First, Gilliland (1993) identified three main categories of perceived procedural characteristics encompassing ten different procedural rules

that impact applicant reactions to the selection process: formal characteristics, explanation, and interpersonal treatment. Formal characteristics include rules such as job relatedness, opportunity to perform, reconsideration opportunity, and consistency of administration. Explanation rules involve the adequacy of organizational communication, and includes such rules as timely feedback, selection information (i.e., justifications being offered for the decisions made during the selection process), and honest communication. Interpersonal treatment rules includes interpersonal effectiveness, two-way communication, and the appropriateness of the questions asked during the selection process. To the extent that a selection procedure violates these rules, applicants are more likely to have negative reactions. With regard to BSJTs, different kinds of branching might violate different justice rules.

Second, person characteristics, which includes demographic characteristics and personality, impact perceptions directly and indirectly through perceived procedural characteristics. For example, Chan (1997) found that Black undergraduates viewed cognitive ability tests less favorably than White undergraduates. Truxillo, Bauer, Campion, and Paronto (2006) found that personality (i.e., neuroticism and agreeableness) predicted applicant perceptions, whereas Viswesvaran and Ones (2004) found that cognitive ability, emotional stability, and conscientiousness all impacted applicant perceptions across contexts. These results suggest that to some degree, individual differences like these affect the reactions of a person to a selection system regardless of the content of that system. Thus, to more fully understand how applicants react to the different kinds of branching present in BSJTs, it is important to measure these antecedents. The outcome of the model is applicant perceptions. Overall applicant reactions to selection procedures depend upon their perceptions of organizational justice, as well as other attitudes (Gilliland, 1993; Gilliland & Hale, 2005; Hausknecht et al., 2004). Organizational justice is comprised of distributive justice, procedural justice, informational justice, and interpersonal justice (Colquitt, 2001; Gilliland, 1993). According to Gilliland and Hale, distributive justice is the perceived fairness of an outcome, procedural justice is the perceived fairness of the procedures used to arrive at an outcome, informational justice is the adequacy of the information provided, and interpersonal justice refers to the

level of respect with which one is treated. Given the purpose of the present study, which focuses upon the design characteristics of a BSJT, the most salient facet of organizational justice is procedural justice.

There is little empirical literature examining applicant reactions to SJTs. In one of the only empirical studies testing this relationship, Chan and Schmitt (1997) examined the subgroup differences among applicant reactions to text-based and video-based SJTs. The authors found that the differences in reactions of Black and White research subjects to SJTs were smaller for video-based SJTs, which was attributable to the increased face-validity of the video-based assessments. Although there appears to be little research on reactions to SJTs beyond this study, Bauer and Truxillo (2006) provide a framework for how applicants would be expected to react to SJTs and assert that five of Gilliland's (1993) procedural justice rules are relevant to SJTs: job relatedness, consistency of administration, feedback, opportunity to perform, and two-way communication. The authors state that the characteristics of SJTs have implications for applicant reactions along these five dimensions. Specifically, SJTs are generally seen as job-related and would presumably be viewed positively in that regard. SJTs are also likely to be rated highly in terms of consistency of administration, at least when compared to less structured assessments (e.g., unstructured interviews). SJTs that provide quicker feedback, such as those scored using multiple choice options, are likely to be viewed more favorably than those that provide slower feedback, such as open-ended questions requiring human assessors. Conversely, SJTs with multiple-choice response options are likely to be seen as less favorable in terms of opportunity to perform. Finally, SJTs with formats that require more interaction with test administrators are likely to be viewed favorably in terms of two-way communication. Unfortunately, these propositions have not been empirically vetted.

There is very limited research on the impact of branching on applicant reactions. Kanning et al. (2006) found that video-based BSJTs were viewed more favorably compared to SJTs in other formats, specifically when compared to text-based BSJTs and video-based SJTs. Aside from this study, there does not appear to be any additional studies examining this effect. However, given Bauer and Truxillo's (2006) model of applicant reactions to SJTs detailed above, it would be reasonable to assert that to the degree

that BSJTs exhibit the salient procedural justice characteristics, they would be more likely to elicit positive reactions.

Although the empirical literature regarding applicant reactions to BSJTs is limited, there are related literatures that are informative. Despite there not being a research literature on branching specifically, there is similar work in several related areas: assessment centers, CATs, and semi-structured interviews. First, assessment centers are relevant because SJTs can be considered a low-fidelity simulation (Motowidlo et al., 1990), which is typically contrasted with high-fidelity simulations, a common component of assessment centers. The term *fidelity* in this context generally refers to the extent that a simulation mirrors actual work conditions. Assessment centers are considered high-fidelity because they require the respondent to act out their responses, such as via role-playing, much like they would in a real-world setting. In contrast, SJTs are generally comprised of text or video-based questions that require only responding to a static prompt by selecting an answer or describing their response in words. Given that BSJTs are at least partially constructed to mirror the impact of a respondent's choices by changing the subsequent scenarios more closely simulates the real-world consequence of a person's behavior, BSJTs are likely to be higher in fidelity than traditional SJTs, although still well short of assessment centers. Second, computer adaptive testing is relevant because of its surface similarities to BSJTs: both adapt dynamically to test-takers' responses. Finally, semi-structured interviews are relevant in that both forms of measurement have some basic structure, but the answers given by the respondent impact subsequent questions.

Perhaps unsurprisingly, there is little research on applicant reactions to assessment centers, CATs, and semi-structured interviews. A study conducted by Macan and colleagues (1994) had subjects in a manufacturing setting complete an assessment center that consisted of a productivity improvement group discussion and a hands-on assembly group, intended to measures skills required to be a good team member, solve problems, and meet production standards (Macan et al., 1994). Before the subjects were notified of their scores on the assessment center, they were given a reactions measure. The authors found

a significant relationship between applicants' perceived control and their assessment center performance ($r = .17$, $p < .05$). They also found that applicant perceptions of the assessment center positively predicted job acceptance intentions. However, the authors point out that these reactions are somewhat mitigated when preexisting impressions of the selection process, job, and organization are taken into account.

Tonidandel et al. (2002) explored how features of CATs impacted 162 test-takers' reactions. The independent variables in this study were the difficulty of the initial item and the difficulty of subsequent items. In the initial item condition, the difficulty of the initial item was manipulated such that it was either easy, moderate, or hard. In the subsequent condition, all items after the first item were constrained such that the probability of getting them right was either 30%, 50%, or 70% based on the current estimated skill level. The authors then measured test performance, the actual difficulty level of the test, performance attributions (i.e., luck, effort, task difficulty, and ability), feedback acceptance, perceived fairness, self-efficacy, satisfaction with the test, test anxiety, and test motivation. Objective test difficulty predicted perceived performance, while perceived performance fully mediated the relationship between objective test difficulty and motivation. Perceptions of performance fully mediated the relationship between objective test difficulty and satisfaction with the test. Self-efficacy moderated the relationship between perceived performance and attributions of performance to the task. Those high in self-efficacy were more likely to attribute their performance to the task when they performed poorly but were less likely to do this when they performed well. For those low in self-efficacy, the relationship between perceived performance and task attribution was positive. Subjects who perceived their performance as poor accepted negative feedback better than positive feedback, while the opposite was true for subjects who performed well. Finally, for subjects who accepted their feedback, the relationship between test performance and post-feedback fairness perceptions was positive for people who perceived their test performance as good. This relationship was negative for those who thought they performed poorly.

These results have an important implication for BSJTs. As the authors point out, because most CATs are designed such that the probability of answering a question incorrectly is almost always 50%,

then in comparison to a non-adaptive test, high ability test-takers are likely to get more items wrong. Thus, high ability test-takers are more likely to have an incongruity between their perceived performance and their actual performance. This study suggests that for those high ability individuals, perceptions of fairness are likely to be more negative than for low ability test-takers. For BSJTs, the implication is that the structure and difficulty of the assessment must be carefully considered as these attributes are likely to have an impact on test-taker reactions.

Roch et al. (2014) examined how the scoring for different selection methods impacted test-taker motivation. The authors had undergraduates complete an objectively-scored cognitive ability test, a semi-structured interview, and written role play, both of which were scored using rating scales. Participants also completed measures of motivation, procedural justice, perceived influence, and perceived performance. Across two studies, results showed the high levels of motivation for the semi-structured interview, particularly compared to the cognitive ability test. However, the semi-structured interview was also rated lowest in terms of procedural justice. Schleicher et al. (2006) studied the relationship between opportunity to perform and fairness judgements across several different selection methods, including semi-structured interviews. The researchers found that opportunity to perform was positively related to perceptions of procedural fairness, and that relative to other assessments (including a structured interview), that unstructured interviews provide a better opportunity to perform. Taken together, the results of these studies indicated that assessments that adapt with the test-taker can have positive effects on important outcomes.

By considering existing literature on SJTs, applicant reactions, assessment centers, CAT, and semi-structured interviews alongside the opinions of practitioner experts, it will be possible to better understand what practitioners are trying to accomplish when they utilize BSJTs. Additionally, these literatures provide a framework to situate these impacts empirically. Specifically, it allows for the testing of the assumption that BSJTs have a positive impact on applicant reactions generally, and applicant perceptions specifically. Thus, the purpose of this dissertation is twofold. First, via a qualitative pilot

study and Study 1, I develop a theory of applicant reactions to BSJTs focusing upon the mediating psychological state changes anticipated by both BSJT practitioners and the extant literature. Second, in Study 2, I quantitatively test this theory.

CHAPTER 2
PILOT STUDY

To generate baseline information about current BSJT intentions and use on which to base a more complete qualitative study, a series of structured interviews was conducted.

Method

Sample. Initially, four SMEs were contacted. Two of these individuals were contacted because they were known by the author to work on BSJTs, whereas the other two were contacted because of their publications in the area of BSJTs. Of the four, three agreed to be interviewed, and the fourth did not reply to my email. A snowball method was utilized to locate other potential SMEs; at the conclusion of each of the three initial interviews, each SME was asked if they knew of anyone else conducting research on or utilizing BSJTs in practice. These recommended SMEs were then contacted, and if that person responded and was interviewed, they too were asked if they knew of any additional BSJT SMEs. Ultimately, nine additional SMEs were contacted. Of the nine, one email came back as undeliverable, two did not respond, two responded but did not feel qualified to be interviewed, and four responded and were interviewed. A total of seven SMEs were interviewed for the pilot study, all PhDs in the field of Industrial-Organizational (I/O) Psychology. One SME was employed by a university, while the remaining six were employed by three different private organizations.

Materials. Interview questions were developed to gain a better understanding of BSJTs in terms of their purpose, mechanics, and psychometric properties. The interview questions covered topics such as why BSJTs are utilized, the different kinds of branching that are utilized, how BSJTs are scored, how the psychometric properties of BSJTs are assessed, their cognitive load relative to traditional SJTs, adverse impact, and the different kinds of constructs that have been assessed using BSJTs. See Appendix A for the complete list of questions.

Procedure. After each SME agreed to be interviewed, and a time to meet was set, the SME was sent the interview questions so they could review them in advance of the meeting. Each meeting lasted approximately one hour, with the author taking extensive notes during each interview. In addition to the

interview questions, the SMEs were asked if they had any documents or presentations they could share. Three of the SMEs were able to share additional information. One provided beta test content, while the other two provided slide decks and notes related to BSJTs. After each interview, the author reviewed the notes and put them into a more narrative form, documented any commonalities with past interviews, and reviewed any additional materials that were provided.

Results

Why branch? The most frequently cited reason for utilizing branching was to enhance the test-taker experience, with SMEs commenting that "candidates seem to like branching" and "anecdotal evidence suggests that people think the assessment is more engaging." Additionally, SMEs indicated that branching more appropriately mimics the kind of non-linearity that would occur in real life, increasing content validity. Specifically, they stated that "decisions have consequences," and branching realistically captures the way decisions in real life impact subsequent events. Relatedly, branching was seen as a way to make the story behind the SJT more coherent to test-takers by having the scenarios directly connected to each other narratively. SMEs also mentioned enhanced test security as a benefit of branching as well as the ability to measure constructs more deeply compared to traditional SJTs.

How does the branching work? One SME explained branching in terms of the kind of content that is impacted by the branching: radical or incidental. In this view, radical content is the content that assesses the construct of interest. Changing this content changes the construct that is being measured and the parameters that have been estimated (e.g., trait level, effectiveness of the response options). Incidental content is unrelated to the construct, and changing this should (at least theoretically) not change the construct being measured or the psychometric parameters. Parallel items, or branches, can be created by changing incidental content but leaving the radical content alone. According to the SME, "the branches aren't easier or harder, they just fit the narrative better. The alternative is to make the test easier or harder based on candidate choices, but then you run into problems of someone getting stuck in a 'low' branch

without the possibility to redeem themselves - there has to be some mobility built into the SJT to prevent this."

The following example shared by one of the SMEs illustrates incidental content branching. A scenario measuring conflict management explains that the test-taker is working on a project with two other co-workers, and that the test-taker's role is to help plan and coordinate the project, with an explanation of each co-workers' strengths and weaknesses. The test-taker is then asked to assign project tasks to the two co-workers. Based on the assignments, the test branches to three different scenarios. The three branches (the scenario and the response options) are nearly identical, varying only in one portion of the scenario. Each branch begins with one of the coworkers explaining how the work is progressing as well as the coworker complaining about the other coworker. The explanation of that progress, as well as the complaint, both narratively follow the response chosen for the previous item. Additionally, the explanation and complaint were purposely written to be brief (i.e., only two sentences) and general so as not to impact the actual construct being measured. Since the purpose of the scenario is simply to set up a conflict that the test-taker must manage, the rest of the scenario and the response options do not need to differ among the three branches.

Another SME made the point that BSJTs are labelled differently depending on the client and the complexity of the content. For example, for one client, the assessment was simply called a "branched SJT," whereas for another it was called "fact finding sequences." In a fact finding sequence the respondent is given the opportunity to talk to certain people or look at certain documents. According to the same SME, "if someone doesn't go down the 'right' path, you can eventually force it by having some sort of transition scenario." In this way, BSJTs can appear similar to certain assessment center exercises but with the information needed to answer the scenarios presented through characters and/or short documents, rather than real life actors. For example, the scenario can present a situation, and ask, "What are the critical things that need to be done?" The possible responses might be to talk to the boss, read relevant materials, talk to an assistant, call a meeting, or collect information. One or more of these options

might be a dead end, requiring the test-taker to go back to the original item and answer again. To be able to score a test-taker on certain competencies, the test might in this way force them to pick an answer that can be scored.

Additionally, not all questions have to be scored, multiple item types can appear in the test, and items may even measure multiple competencies. Regarding scoring, a respondent can be asked to rate the effectiveness of each response, as well as how they would actually respond, with the latter response being scored and used to determine how the narrative moves forward. One SME stated that, "multiple competencies can be assessed in a single question, but that can make it harder to make sure someone sees all items for a particular competency unless the branching forces it. If the test doesn't force it - that can result in some competencies having more data than others."

One issue that must be addressed during design is whether the different branches eventually converge. There may be branching, but typically the test is designed to converge back to a single point. The number of branches and which items result in a new branch may be dictated by the client. It may be that, realistically, each response could lead to dramatically different events, which would need to be captured by the branches. Overall, realism is a key driver of the decision to branch.

Does branching create any limitations? Two SMEs indicated that assessment equivalence is a major limitation. According to one SME, "branching means everyone gets a different assessment." As one SME pointed out, this is "more of a challenge for an assessment context" compared to a developmental context. For selection, this might mean that the measures aren't equivalent. The purpose of the assessment also impacts how the scenario ends. As one SME put it, "For developmental BSJTs, it may be fine for the scenario to end early as the result of good or poor choices, since this would allow the scenario to lead to the natural consequences. For selection, you wouldn't want a scenario to end in a disaster."

Two SMEs also stated that the technology behind the BSJT was a limitation. For example, compared to a live roleplay, a BSJT contains a finite number of response options, so the response format

is more organic than a traditional SJT but still not as open-ended as a live interaction. One SME emphasized the practical limitations given the particular media used to administer the BSJT (i.e., video vs. animation vs. text-based). For example, a video-based SJT requires substantial resources to create relative to a paper and pencil SJT (Chan & Schmitt, 1997). A video-based BSJT would require even more resources if branches require new video to be created. This could significantly add to the cost of creating the BSJT. The SME stated that "for a live action BSJT, it is very expensive to create a scenario compared to an animated scenario, which is more expensive than a simple text item."

Other limitations mentioned by SMEs were related to test security and retesting. For BSJTs where the branches are more or less parallel (i.e., the incidental content changes but not the radical content), test security is more concerning, because content may be compromised more easily. There is also little guidance on how branching impacts retesting, so it is difficult to base retesting policies on a common, widely accepted standard.

How is the BSJT scored? The SMEs described multiple ways to score a BSJT. One SME described a process wherein SMEs estimate the trait level of each response option along a certain competency, as well as the effectiveness. These values are used to form a composite. Another SME indicated they worked on a BSJT where the composites were created with mean centering based on a stretching algorithm (based on McDaniel, Psotka, Legree, Yost, & Weekley, 2011) that adjusts for elevation and scatter. According to the SME, "this is particularly helpful if there is a lot of range restriction." Another SME used simple effectiveness scoring on a seven-point scale, asking test-takers to evaluate how effective each response was given the scenario.

The SMEs also mentioned other creative ways that a BSJT could be scored. One SME had heard of, but not had any actual experience with, the use of Bayesian methods to score BSJTs. Another SME suggested that because SMEs generating items do not always agree on which response option is the "right" response option, several methods could be used within the same test. For example, the test-taker could choose what they would do in a particular situation, list the response options in order of priority,

indicate how soon they would do each response option, or rate the criticality of each piece of information presented in the scenario. A distance score could then be created for each response, taking into account mean centering and stretching. However, the SME indicated that using multiple scoring within a single BSJT could present some scaling problems, since some response options might be presented on one scale, while other response options might be on another. One SME stated that they had hoped to use item response theory (IRT) to determine response option weights but was unable to attain a sufficient sample size to obtain stable parameters. In lieu of IRT scoring, deviation scores were used.

One SME commented that competencies can be measured using a BSJT but that a balance must be struck between test length and candidate experience. According to the SME, "you have to think about how long the assessment could be before it negatively impacts user experience but had to be long enough to be meaningful."

How are the psychometric properties of the test assessed? Regarding reliability, one SME assessed reliability via test-retest reliability, because the test they were utilizing might not have been "construct-homogeneous." Another SME indicated that their organization was still trying to determine the best method for measuring reliability. One of these SMEs remarked that "it gets complicated with regard to reliability and validity" and, along with another SME, stated that Putka, Le, McLoy, and Diaz's (2008) study on assessing reliability was often consulted because it provides information on assessing internal reliability for data "that has a lot of empty cells." Only one SME mentioned content validity, and stated that it was assessed before the test was administered. The same SME also mentioned that they were in the process of collecting evidence for criterion validity.

Is adverse impact (AI) higher or lower compared to traditional SJTs? One SME stated that "research is still out on that." Another SME also cited their BSJT medium (i.e., video) as a reason that AI wasn't much of a concern, because "moving to video usually reduces subgroup differences – not sure about branching, but there is no reason to suspect" that subgroup differences would be an issue. One SME

indicated that the samples they had were not "very diverse," but that the available data indicated that they were similar in AI to assessment centers, stating differences were "about ¼ or ½ of a standard deviation."

Are BSJTs more or less cognitively loaded than traditional SJTs? One SME suggested cognitive load was likely to be low given that most of her organization's BSJTs are video based, and thus do not require much reading, an assertion supported by Chan and Schmitt (1997). Another SME stated that BSJTs are not more cognitively loaded "by definition, but some are because the constructs are more cognitively loaded." That is, it is important to distinguish between the method of measurement and the construct being measured (Arthur & Villado, 2008). Another SME indicated that most of the BSJTs with which he had experience measured soft skills and tended to require little cognitive energy, but that "some of the measures appear to be heavily *g*-loaded." In a traditional SJT, every item is independent and thus the demand on working memory is minimal because one's information store essentially resets after each item. The opposite is true for a BSJT, in which branched items dependent on the previous item. As a test-taker proceeds through the test, the story develops, and there could be additional demands on working memory due to an implied expectation that the test-taker remember details of the story throughout the assessment. Thus, BSJTs may be more *g*-loaded than traditional SJTs. Alternatively, aspects of BSJTs suggest a reduction in cognitive demand on test-takers. Since a narrative is present throughout the test, less of the test can be devoted to the stimulus scenarios because an entirely new scenario does not have to be created for each new item. The developing story provides additional context for each item that does not have to be explicitly detailed in each item, unlike a traditional SJT. Thus, the net effect of these two forces is unknown. Given this, a better understanding of the mediating psychological changes involved in BSJT completion could be developed by empirically exploring the relationship between BSJT design and cognitive load.

Does branching impact the underlying construct being measured? The branching/construct relationship depends on the purpose of the test. One SME stated that if the "meaning represented by the path is important," then the test developer has to determine how to ensure key items are seen, "maybe

through parallel items, or forcing them through the narrative." For example, if an item measures conflict management, and the choices made by the test-taker have consequences that cannot be undone, then the challenge for the test developer is to determine how to best measure conflict management with the different branches. As the SME stated it, "Branching increases realism but presents measurement issues." One SME indicated that different candidates may see information at different points in their respective assessments. A particular item may require certain information to answer presented earlier in the assessment, but test-takers may see it at different times in their assessment.

What constructs have you measured using BSJTs? BSJTs seem to mainly be used to assess soft skills, although hard skills could be assessed. One SME stated, "Because [the BSJT] involves choices, ones that involve problem solving, decision making, and interpersonal skills are ideal. Conflict management is also a good one because the conflict can build or evolve over time instead of just a single item like other SJTS." BSJTs allow test developers to create more complex scenarios. SMEs listed the following additional constructs: judgment and problem solving, relating to others, conflict management, leading others, planning and organizing, knowledge tests, investigative procedures, administrative procedures, working with people, interpersonal skills, and developing others. As one SME put it, "You can assess in a BSJT what you can assess in an SJT, but the BSJT may allow for deeper measurement." However, one SME cautioned that it is important to recognize that various responses to a single question may not tap the same construct. The effectiveness of each response option can measure different constructs. Or, test-takers might translate response options in their own way, and thus may respond to a slightly different interpretation of a question than another test-taker.

What are the important questions that need to be asked/answered regarding BSJTs? There are a number of lingering questions that SMEs communicated, including understanding how changing response options impacts construct validity, the impact of the question order and timing, the optimal way of computing reliability, the impact of mean centering, the ideal balance between the number of measurement points for each competency and overall test length, and the types of questions that can be

appropriately asked ("Can you rank each item in the order that you would do them?" If so, how would they be analyzed?"). One additional SME commented that there are "so many scoring issues" that a great deal of empirical research is needed across the entire domain.

Discussion

The seven interviews yielded a great deal of insight into BSJTs. From these initial interviews, the impetus for the use of BSJTs seems to generally involve enhancing the candidate experience. Several SMEs mentioned, either explicitly or implicitly, that BSJTs provide a more engaging experience via a more coherent narrative and enhanced realism relative to traditional, static SJTs. Research has indeed linked candidate reactions to a number of positive outcomes, including organizational attractiveness (Bauer et al., 1998; Bauer, Truxillo, Sanchez, Craig, Ferrara, & Campion, 2001), intentions toward the organization (Bauer et al., 1998), recommendation intentions (Bauer et al., 2001; Gilliland, 1994; Gilliland, Groth, Baker, Dew, Polly, & Langdon, 2001), decreased litigation likelihood (Bauer et al., 2001), perceptions of fairness, and reapplication intentions (Gilliland et al., 2001). Current research supports the premise that BSJTs result in positive applicant reactions. Kanning et al.'s (2006) study comparing different SJT types found that candidate reactions were greater for an interactive (i.e., branching) SJT compared to a static one. However, Kanning et al.'s findings may be limited in scope due to the nature of the BSJT they utilized. In their study, the BSJTs the researchers used had only 2 response options. One response option took the test-taker down one branch, while the other option led down a second branch. The penultimate items (one for each branch) in turn had two response options, again which had two potential branches. While useful as an initial examination, the BSJTs described by the SMEs are substantially more complex than the one used by Kanning and colleagues. Different, more complex types of branching may have a more positive (or a more negative) impact on applicant reactions. Given the SMEs' assumptions that branching results in improved applicant reactions, it is critical to examine this relationship further.

BSJTs can vary greatly in terms of the kind of content that changes among branches (incidental vs. radical), how they are referred to (e.g., BSJT, fact finding sequence, etc.), and in terms of the number of items within the test that are scored. This variability in format and content makes it somewhat difficult to make more general statements regarding BSJTs. Additionally, as stated previously, it is important to remember that SJTs are a method of measurement, and, as such, issues concerning method must be examined separately from issues concerning specific constructs (Arthur & Villado, 2008). If the branches of one assessment differ in terms of incidental content, while in another assessment the branches differ radically, a comparison of the effects of these two assessments may not be appropriate. This is likely to be true for any two BSJTs, given the number of different ways branching may occur. Further complicating such comparisons is the lack of understanding of the ways that SJTs can branch. For example, there is currently not a model of the various ways BSJTs can branch or the impact of each of these kinds of branching on other properties of the assessment (e.g., applicant reactions, reliability, validity, etc.). A greater understanding of the different ways SJTs can branch would assist researchers in designing studies to explore important theoretical and practical questions. There appears to be little in the way of guidance regarding how to assess reliability or validity, or how BSJTs may impact subgroup differences. Again, even if there was some empirical evidence to rely upon, the question of what kinds of BSJTs the evidence applies to remains. There is some consensus regarding the kinds of constructs measured using BSJTs, with 'soft skills' being the main constructs assessed due to the nature of BSJTs. Again, questions of whether certain kinds of branching hinder or enhance the measurement of certain constructs remain.

This pilot study, while limited in the number of participants, has provided rich, detailed information regarding BSJTs. Specifically, it has provided some insight into the assumptions of users of BSJTs in terms of why they are being used, how they are scored, how their psychometric properties are assessed, and the kinds of constructs they are used to measure. One major theme that has emerged is that at least for the small number of SMEs interviewed for the pilot study, BSJTs are used to enhance applicant reactions. However, this is troubling, given that little is known about the different ways SJTs

can branch, and how different kinds of branching may impact applicant reactions. Clearly, a more in-depth study is needed to determine if the themes that emerged in this pilot study are generalizable, or simply an artifact of the limited number of SMEs interviewed. Thus, Study 1 will expand on the pilot study by using grounded theory to explore BSJT-related applicant reactions in more depth.

.

CHAPTER 3
STUDY 1

The pilot study illuminated a number of important issues and assumptions with regard to why and how BSJTs are constructed. The primary motivations behind utilizing branching were a) enhancing the test-taker experience, b) enhancing test security, and c) allowing for deeper construct measurement compared to SJTs. Branches can differ a little bit or a lot (e.g., incidentally or radically), which is in turn suggested to impact measured constructs. Branching can be driven by a narrative, but it can also be a function of how the test-taker is performing, similar to CATs. Much like SJTs, BSJTs can vary widely in terms of how the items are scored, from simple least effective/most effective response options to the use of Bayesian methods. Potential limitations include comparing scores when test-takers don't necessarily see the same items, the close-ended nature of the response options compared to other measurement methods (i.e., an assessment center), the lack of guidance regarding retesting effects, and the expense associated with test creation, particularly for video-based tests. There are not yet any specialized methods of assessing reliability or validity, and there appears to be no reason to suspect that AI would be substantially different for BSJTs compared to SJTs. Finally, BSJTs can be used to assess both hard and soft skills, though they are most often used to assess soft skills.

One theme that emerged from the pilot study was that there is not "one" way to branch an SJT. Branching can be based on simply furthering a narrative, or it can be driven by how well the test-taker is doing. Some test creators make significant effort to ensure that the different branches are as similar as possible, while others seem less concerned about the comparability of branches. A BSJT can converge at a certain point, forcing test-takers to specific items, or they may allow test takes to follow whatever path they wish. While the pilot study has uncovered some features of BSJTs, there may be others that have yet to be delineated.

That BSJTs enhance the test-taker experience was another major theme across the pilot study interviews. Branching is thought to make the make the assessment more engaging, to more realistically portray how events would unfold in real life given test-taker responses, and build a more coherent test via

a narrative that unfolds throughout the test. However, no empirical evidence was cited to support this assumption. Indeed, as discussed earlier in this paper, very little empirical evidence exists.

The purpose of Study 1 was to build and extend upon the preliminary pilot results to examine how BSJTs are being used by scientists and practitioners to develop a theory of SJT branching. A qualitative study was conducted to better understand the different ways SJTs can branch, what applicant reactions are potentially impacted by branching, and how specific forms of branching are believed to impact applicant reactions.

Method

Data collection for Study 1 was guided by Schonfeld and Mazzola's (2013) approach to grounded theory, a method of qualitative data collection developed by Glaser and Straus (1967) that allows researchers to generate theory through the systematic collection and iterative interpretation of qualitative data (Glaser, 2010). Grounded theory is useful in situations where not much is known about the phenomenon in question (Glaser & Straus, 1967). Several researchers have utilized grounded theory to explore organizational phenomena. For example, Wilhelmy, Kleinmann, Konig, Melchers, and Truxillo (2016), through speaking with interviewers, observing interviews, and reviewing documents related to the interview process, used grounded theory to develop a model of interviewer impression management intentions, behaviors, and outcomes. Likewise, Isabella (1990) conducted in-depth interviews with 40 managers to develop a model of how managers view change events over time, and Wang, Lo, Xu, Wang, and Porfeli (2007) reviewed diaries and conducted interviews with 4 doctoral students to better understand the job search process of recent graduates seeking academic positions. Given the dearth of empirical literature regarding BSJTs, grounded theory provides a useful framework from which to explore the different ways that SJTs might branch and enable the development of a comprehensive model of SJT branching.

O'Reilly, Paper, and Marx (2012) describe the five main tenants that guide good grounded theory research: the constant comparative method, theoretical coding, theoretical sampling, theoretical

saturation, and theoretical sensitivity. While each is described only briefly here, a more in-depth explanation of grounded theory and its tenants can be found in Glaser and Straus (1967), O'Reilly et al. (2012), and Strauss and Corbin (2012). The first tenant is constant comparison, in which data collection and data analysis occur iteratively, with each one informing the other. As the data are collected and analyzed, the results of that analysis are used to adapt and guide additional data collection. As data are being collected, they are theoretically coded into groups and categories to discern the theoretical underpinnings of phenomena under investigation. Theoretical sampling is sampling guided by the data, which lead to further refinement of the concepts uncovered by those data. This process of data collection and analysis continues until the process ceases to generate novel information. It is at this point that theoretical saturation (or category saturation) has occurred, which signals to the researcher that sufficient data have been collected. Finally, theoretical sensitivity refers to a researcher's ability to give meaning to data, separating relevant data from irrelevant and properly distinguishing between the two. It is these five principles that guide the following methods.

The current study, although guided by Glaser and Straus' (1967) original conceptualization of grounded theory, instead took Schonfeld and Mazzola's (2013) adaptation of the technique to accommodate the practical constraints of empirical research. According to Glaser and Straus, the themes and ideas that emerge from data in grounded theory should be free of researcher preconceptions, but Schonfeld and Mazzola described how this is a difficult goal to attain, because most researchers know a great deal regarding important concepts and theories related to the phenomena they are investigating before collecting any data. To reconcile grounded theory with this practical consideration, Schonfeld and Mazzola recommended that authors try to acknowledge the ideas and biases that may impact their investigation and do their best to keep them from becoming a self-fulfilling prophecy. With regard to the present study, steps were taken to ensure that the ideas, concepts, and themes that emerged from the study were as free from bias as possible. During the interview process, I made an effort to not put words in the interviewee's mouth or ask leading questions so that the data emerged as naturally as possible.

Additionally, during the coding process, the appropriateness of the categories and codings were questioned often, and new, more specific questions were asked throughout the process to revise my understanding based upon data collected to that point.

Data were collected from several different sources, including semi-structured interviews, archival documents, and actual BSJTs. To locate BSJT subject matter experts (SMEs) to interview, first authors on every published paper since the year 2000 that contained situational judgment test as a keyword, as well as first authors on conference presentations, symposiums, and posters were contacted via email and asked to submit to an interview regarding their knowledge of BSJTs. The year 2000 was chosen as a practical cut-off, as any authors who have not published on the topic since then are unlikely to be up-to-date on such a recent innovation as BSJTs. This approach yielded an initial list of 261 potential SMEs. Contact information for 11 SMEs could not be located, and 250 email invitations were sent. Additionally, an announcement was posted on two electronic mailing lists, including those managed by the Academy of Management's Organizational Behavior and Human Resources Divisions, which resulted in the addition of two further SMEs. A tweet was also sent with the hashtags #iopsych and #siop. Posts were made on the Society of Industrial-Organizational Psychology's and Talent Selection and Assessment's LinkedIn pages. None of these methods yielded additional contacts.

The announcement included the contact information for the author, a description of the purpose of the study, and a request for 30 minutes of the SME's time. While the interview utilized a predetermined set of questions, these questions only served as the foundation of the interview and other questions were asked when appropriate. A snowball technique was used to locate other potential SMEs which yielded 14 additional contacts. SMEs were also asked to share materials related to BSJTs, including beta and live assessments, descriptions of past, present, and future assessments, and any documents that might be relevant. Three dissertations, a book chapter, and information on a research study were provided in response to this request. After all interviews were completed, the interviews were transcribed and, along with the additional materials gathered, reviewed and coded. Interviews were

conducted until theoretical saturation was achieved. Specifically, for the first interview, several questions were determined prior to the interview, but additional questions were asked to probe certain responses. For example, in one of the early interviews, one of the first interview questions was "Can you tell me about the mechanics of the test?" In subsequent interviews, additional follow up questions were added, such as if the SME could walk me through the tests, how the branching works, what triggers the branching, how the content of the branches differ, and if the constructs of the branches are the same or different. As the interview process proceeded, the interview questions were refined, with questions being dropped, modified, or added.

A total of 24 interviews were conducted. Of these SJT experts, six stated they were not BSJT experts, resulting in a final sample size of 18. These interviews were recorded using TapeaCall, an iPhone app. The SMEs suggested eight additional resources, including internet links, an applicant reactions survey, three dissertations, and a book chapter, but only one dissertation and the book chapter were found to be codable (i.e., included information specific to BSJTs). Thus, 18 interviews, 1 dissertation, and 1 book chapter were coded.

Sample. A total of 18 SMEs were interviewed, 12 male and 6 female. The SMEs came from a number of industries, including Research and Development, Consulting, Professional Services, and academia. One SME had a master's degree, and all the remaining SMEs had PhDs. In my judgment, theoretical saturation was attained around interview number 12 in that I was no longer hearing new information in relation to my research questions in response to either my original questions or my revised questions given earlier interviews. However, I was concerned that, given the limited number of SMEs, certain populations of SMEs might have been missing from the first 12 interviews. Thus, I interviewed all BSJT experts identified to ensure that there were no meaningful gaps in the data.

Analysis

Traditional grounded theory methods were adapted for this study to create a method similar to that used by O'Reilly (2010). These adaptations are the addition of interview recordings and member

checks. To ensure accuracy, all interviews were recorded and transcribed. The present author conducted all of the interviews. In keeping with the constant comparison method, interview questions were added, dropped, or revised in order to incorporate new information into the interview process. Another graduate student with expertise in I/O psychology was trained in grounded theory methods and also reviewed the interview transcripts to provide a second set of codes. To address RQ1 through RQ4, a multi-step process was used, keeping in mind the tenants of grounded theory described above. First, the two raters independently reviewed the interview transcripts line-by-line and developed codes independently. Before this coding took place, the second coder was trained by providing specific instructions on how to consider their own knowledge and potential biases and to keep an open mind with regard to the emerging concepts and ideas. Second, the codes were revised during a collaborative review process (Kreiner, Hollensbe, & Sheep, 2009) and were entered into a spreadsheet. All other materials collected during the interview process (e.g., archival documents) were coded as well. Third, the coders met multiple times to discuss the codes, with particular attention toward evolving themes, in order to resolve any coding discrepancies but also to continue to apply the constant comparison method and further refine the data being reviewed. Finally, the concepts and categories were reviewed to determine in what ways they were related and distinct. Once the categories and concepts were finalized, a member check was done whereby the models (detailed below) were sent to the interviewees to ensure that they are meaningful and based on the data. A document summarizing the findings of Study 1, including the theoretical models generated from Study 1 was sent to all 18 SMEs. Nine provided feedback, including suggested name changes to the various branching features, comments on the paths within the model, and suggestions for ways to clarify the model and branches. No SME suggested major changes in the descriptions or models. Thus, the member check confirmed the accuracy of the results to be described next.

Results

RQ1. Through the interviews and resource reviews, four key branching features were identified: *contingency, convergence, parallelism, and looping*. First, BSJTs progress between questions based upon

one of two contingencies: narrative progression alone or narrative progression plus test performance. The basic purpose of branching in an SJT is to tell a sort of story with the assessment. As one SME put it, a BSJT allows the test creator to "create a kind of an overarching story or a narrative or at least an overarching contextual description that can accommodate the entire SJT." Thus, a narrative flow is a basic characteristic of all BSJTs. However, some BSJTs also branch based on how well or poorly the test-taker is performing on the assessment (e.g., test-takers choosing more effective responses see different items than other test-takers). The test-taker's performance could result in the test-taker seeing an entirely different scenario, or different options for the same scenario. With the latter version, the options the test-taker sees are based on their score on the previous item. If they chose a more effective response, then the subsequent response options reflect that. If the test-taker chose a less effective response, the response options would reflect that too. One SME stated that "if they're choosing the best responses in the first couple of scenes, then [for the next scenario] they would be given the four higher response options to choose from."

Second, parallelism refers to the degree to which the content of the different branches measure the same construct. A BSJT where two test-takers proceed down two different branches but experience the same measurement opportunities is considered parallel. A BSJT where different branches result in different measurement opportunities is considered non-parallel. BSJTs can vary greatly in terms of parallelism. Branches can differ only slightly, with the name of the person changed between branches so that the scenario properly follows the previous response, or the scenario can be entirely different while also assessing a different construct. Most SMEs prefer to stay somewhere between these two extremes, allowing the branches to differ while considering the measurement equivalence of the overall test. As one SME stated, "conceptually the way we always construct these is it kind of continues to unfold, but everybody who takes the assessment gets a large majority of the exact same questions or the same interactions."

Third, convergence is a key feature of BSJTs. Converging the BSJT at certain points allows the test creator a way of limiting the scope assessment in terms of both test creation and construct measurement. That is, convergence allows test creators to keep the assessment at a reasonable length in terms of number of branches, but also to ensure that all test-takers see the same information. Weekley et al. (2015) discuss a type of BSJT where the branches never converge. That is, each response leads to a new branch, and so on and so forth, expanding the BSJT exponentially as one proceeds through the assessment. In practice, BSJTs are generally not constructed this way because such branching is highly impractical. Thus, the primary distinguishing characteristic is not whether the different branches converge, but rather how and when do they merge. In this regard, there are two primary ways BSJTs converge: parent items and forced convergence. Parent items are base scenarios that all test-takers see. Some or all response options from a parent item will lead to a distinct scenario or animated/video-based cutscene. This branching can continue, or it can end, leading to another parent item. As one SME described parent items, "you would have question one and then question two would be based on the response to that. So, [there are] four different versions of question two, but then when you come to question three those multiple versions come back to a single question." Forced convergence occurs when a test-taker is forced by the test to a certain point, typically so that they encounter vital information or a vital measurement point. For example, consider a scenario in which two test-takers progress down two separate branches. In one branch, the test-taker encounters Joe, who conveys information necessary to answer questions later in the BSJT. In the second branch, the test-taker does not encounter Joe and consequently does not get the relevant information. To ensure that the second test-taker is not disadvantaged by their choices, at the end of their branch, they are forced to interact with Joe, either through a scenario or a cutscene, to "push them back to essential information."

Fourth, looping occurs when a test-taker revisits a portion of the test, which can be by choice or required by the test designer. Looping occurs along a continuum from no looping to the ability to revisit any portion of the test at any time. Within this continuum, there are two main types of looping: choice and

forced. Choice looping occurs when the candidate is given the choice to go back in the assessment or to continue forward to new scenarios. This choice might be provided because a candidate is being given the opportunity to explore alternate paths earlier in the assessment or as part of the narrative. In an example given by a SME, test-takers could repeat content to get a better score "if the candidate asks them do you need something additional to that then they go off on a little loop, so they get the main points of, like a point for offering an up-sell, and then the additional looping" to follow the choice that the test-taker made. Forced looping refers to the candidate being pushed back to a certain point in the test. This feature is often used to ensure that a test-taker encounters certain information or scenarios that they may have missed because of the path they went down. Note, in neither choice looping or forced looping are candidates able to alter their initial responses. They are simply able to go down additional paths. Within both choice and forced looping, some test creators allow test-takers to perform a "save" to regain points lost due to a mistake. That is, saving is the ability of a test-taker to go back and actually change their initial answer. While this results in them going down a different path, it also allows them to change their score on the assessment. Consider the not uncommon scenario of a professor allowing students to change their answers on an already graded test for partial credit. Saving is similar – it allows test-takers to change their answers, most likely for partial credit.

RQ 2. In identifying the diverse ways that branching impacts applicant reactions, several themes emerged from the data suggesting that BSJTs impact applicant reactions via perceived procedure characteristics and applicant perceptions. The three most salient perceived procedure characteristics are face validity, consistency of administration, and opportunity to perform. BSJTs are likely to be viewed as face valid because of the way that they realistically capture the consequences of decisions and mimic the demands of the actual job. As one SME put it, a BSJT "feels and looks like a day in the life, so there is higher face validity than, say, taking a generic personality assessment." BSJTs also impact a test-takers perceived opportunity to perform because they adapt to the choices of the test-taker which gives the test-taker the sense that they can demonstrate their true skills. One SME noted that, "if I know that the test is,

kind of, growing and working with me, it also gives me a chance to show my true abilities." The final procedure characteristic through which BSJTs impact applicant perceptions is via the test's perceived consistency of administration. However, the impact of consistency of administration on applicant reactions was described as positive or negative. On the positive side, a test that is customized to the test-taker might make the test-taker feel that they are getting an assessment tailored just to them. As one SME said, branching creates a "wow factor" by creating a unique experience for each candidate. However, test-takers might not like the fact that the test taking experience is not standardized and that they might be taking a somewhat different test than someone else. As another SME stated, "different people will have different psychometric experiences and be given different opportunities to respond to different segments of the branching structure," and this experience might negatively impact applicant reactions.

The perceived procedure characteristics that BSJTs impact in turn impact applicant perceptions: test motivation, attitude toward the test, and perceptions of procedural justice. BSJTs increase test motivation because the story-like nature of the test makes candidates feel more immersed in, and more engaged with, the test. As one SME put it, a BSJT is likely to be "more immersive for the candidate to feel like they're inside a story rather than, you know, a bunch of individual test questions." This feeling of immersion and engagement leads to increased test-taking motivation. BSJTs are likely to be viewed positively as a selection measure because of the interactive feel that the assessment creates. One SME noted that, "the less it feels like a test, the more fun it is." Finally, BSJTs are likely to enhance application perceptions of procedural justice. A test that adapts to test-taker responses will make test-takers feel like the test itself is fair, because it driven by their own choices and not some predetermined path set by the test creator.

RQ3. Figure 2 depicts the first theoretical model of branching that emerged from coding the interviews, archival documents, and empirical literature. It depicts the impact of the use of branching in general in comparison to an SJT that does not have branching. Figure 3 depicts the second theoretical model of branching that emerged, which concerns the impact of branching design. Narrative branching is

the focus of Figure 2 and absent from Figure 3 because it is the most fundamental feature of all BSJTs. The presence of narrative is variable only when comparing SJTs to BSJTs and not when modeling across types of BSJTs.

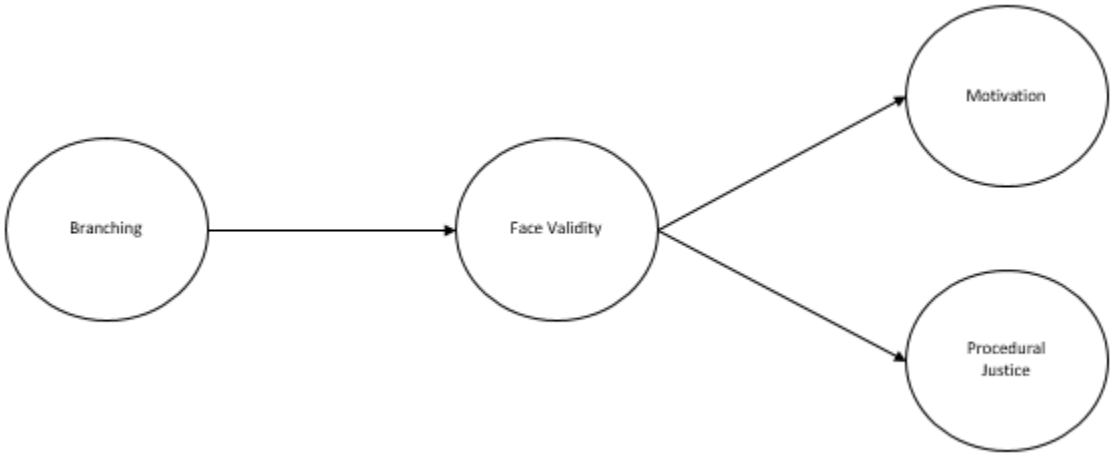

Figure 2. Theoretical impacts of the general use of branching in SJTs.

Comparing SJTs and BSJTs. As shown in Figure 2, the use of branching is likely to impact the apparent face validity of the BSJT. Modern BSJTs always convey a narrative that adapts itself according to test-taker decisions. This adaptation is likely to be seen as more reflective of actual jobs because in real life, decisions have consequences which lead to more decisions, and so on. In a BSJT, this dependency is mimicked by branching. The enhanced face validity of narrative branching is theorized to lead to increased test motivation based upon Hausknecht et al. (2004), who found that tests perceived as face valid led to increased test motivation in both authentic (ρ = .31) and hypothetical contexts (ρ = .41). Test-takers will see the story unfolding via the narrative branching as representative of actual situations they

may face on the job. This perception is likely to result in an enhanced desire to score well on the assessment for BSJTs in comparison to SJTs (Hausknecht et al., 2004).

Comparing Across BSJT Designs. As shown in Figure 3, three BSJT features differ in common BSJT designs and have distinct theoretical effects on reactions. First, an assessment that branches based on performance is similar to how CATs adapt to test-taker performance in that the construct or skill level of the test-taker drives the selection of future test questions. Tonidandel et al. (2002) found that for CATs, the perceived ability level of the test could create an incongruity between perceived performance and actual performance. Specifically, high ability test-takers noted that despite their ability level, the test was still difficult, which occurs due to the standard mechanics of CATs. In such a situation, the consistency of administration is likely to be a salient perceived procedure characteristic. A BSJT that branches contingent upon performance, that adapts as someone does better or worse, is likely to cause people to attend to the fact that the test is different among test-takers. However, as stated before, the lack of consistency of administration could have both a positive or negative impact, as some test-takers might see the branching as a compelling innovation, whereas others might be concerned that they are not being measured the same way as other test-takers. These two diverging reactions might then have an impact on test-taker attitude toward the test and on perceptions of procedural justice (Hausknecht et al., 2004; Tonidandel & Quinones, 2000) depending on their subjective evaluation of the lack of consistency of administration.

Second, parallelism concerns the similarity among the different test forms that test-takers encounter. Branches can vary greatly in the content they present to test-takers. In some instances, two branches may be virtually identical with only a few nouns or other minor bits of language changed so that the branch is consistent with the previous response. In other instances, the two branches could be completely different, with entirely different scenarios, characters, and outcomes. Additionally, different branches need not necessarily measure the same construct. Thus, there are many ways that branches can differ, varying by degree. Much like performance-based branching, parallelism is likely to impact test-

taker perceptions of the consistency of administration of the test. To the extent that the constructs measured and the scenarios encountered are the same or similar, test-takers are likely to have more positive perceptions of consistency, which is in turn likely to lead to a more positive attitude toward the test and more positive perceptions of procedural justice (Hausknecht et al., 2004; Tonidandel & Quinones, 2000).

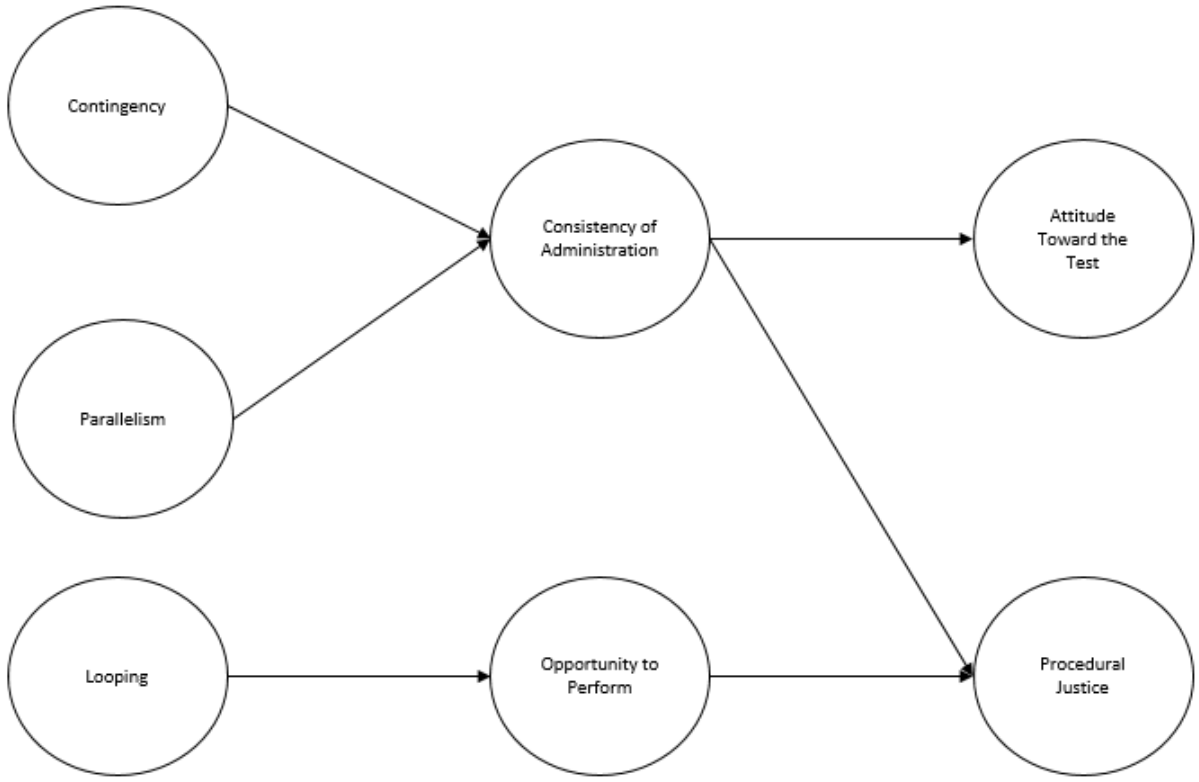

Figure 3. Theoretical model of the impacts of branching features.

Third, looping occurs when a test-taker, either by choice or by force, revisits a portion of the assessment. The effect of looping is that test-takers can explore additional choices or even undo previous choices. Test-takers can take different branches and see how those different choices play out or to obtain

different information. The more looping is permitted, the more the test-taker is likely to perceive the test as providing opportunities to perform. This enhanced perception of opportunity to perform is likely to lead to more positive perceptions of procedural justice (Hausknecht et al., 2004).

Omitted Constructs. Although identified as BSJT features when addressing RQ2, convergence, forced versus choice looping, and saving during loops were not included in the models outlined above for parsimony and simplicity in testing. Specifically, convergence is used by test creators to ensure that all test-takers see certain content or to constrain the length of the assessment via parent items, forcing test-takers to certain scenarios or other mechanics. Convergence has the effect of a) constraining the number of scenarios that must be written by test creators and/or the number of scenarios encountered by test-takers, and b) ensuring that all test-takers have access to the same information. Although this design feature is important from a test-creator perspective, both approaches result in similar experiences from a test-taker perspective and are therefore unlikely to substantially impact applicant reactions. Similarly, the use of forced versus choice looping and the ability to "save" during a loop are both design decisions made subsequent to the decision to include looping itself. Thus, although these are important design features, they are secondary to the looping mechanism and likely to have much smaller effects. Accordingly, their impact is left to be explored in future research.

RQ4. Although this study was intended to gather information on how applicant reactions are measured uniquely for branching, SMEs generally did not gather unique data about branching and how it impacted applicant reactions. In some instances, applicant reactions were collected on the test battery or the BSJT in terms of its performance as an SJT, but this study did not address RQ4. Thus, it was not possible to answer RQ4 from the data collected from this study. However, given that the themes that emerged from research questions one through three concern constructs and variables (i.e., perceived procedure characteristics and applicant perceptions) prevalent in the extant literature, measures from that literature should be sufficient to test the models depicted in Figures 2 and 3.

Discussion

The results of Study 1 provide some of the first theoretical evidence regarding important questions related to BSJTs. Specifically, the 18 interviews with BSJT SMEs provided information on a) the different ways SJTs branch, b) the potential impact of this branching on applicant reactions, and c) how branching is used to impact applicant reactions. The results of Study 1 indicated that there are four main features of branching: contingency, parallelism, convergence, and looping. BSJTs can be driven solely by narrative progression, or branching can be contingent on test-taker performance. The different branches of a BSJT can be created so that the constructs measured and the scenarios encountered are parallel and virtually the same, save for some strategically altered wording, or the content between branches can be dramatically different. Branches can converge through parent items which act as a way to bring all test-takers back to a common item, or test-takers can be forced to certain items to ensure that they encounter certain, vital content. Finally, test-takers can be allowed to loop back to previous content, or they may only be able to follow a linear path through the test.

While the SMEs discussed a wide array of applicant reactions, several themes emerged from the interviews. First, the perceived procedure characteristics of face validity, consistency of administration, and opportunity to perform emerged as important constructs, as did the applicant perceptions of test motivation, perceptions of procedural justice, and attitude toward the test. Further, by coding and analyzing the interviews, and synthesizing these findings with the extant literature, two theoretical models of the impact of branching on applicant reactions emerged. The first model (see Figure 2) detailed the impact of branching in general on applicant reactions. Specifically, the model stated that branching in general is likely to impact both motivation and perceived procedural justice, but these relationships will be mediated by face validity. At their core, BSJTs tell a story over the course of the assessment, and this story is at least partially guided by the test-taker's choices. Such an assessment is likely to enhance test-taker perceptions of face validity as the story told through the BSJT mirrors the way events would unfold in real life. This increased face validity should lead to enhanced test-taker motivation perceptions of fairness.

In addition to a model of the effects of branching in general, another model (see Figure 3) was developed to describe the effects of each branching feature on the relevant applicant reactions. Contingency branching was theorized to impact perceived procedural justice and test attitude via consistency of administration. Parallel branching was also theorized to impact perceived procedural justice and test attitude via consistency of administration. Looping was theorized to impact perceived procedural justice via the opportunity to perform. For the sake of parsimony and simplicity in testing, several constructs were omitted from the theoretical models. While an important branching and test design feature, convergence is unlikely to have a large impact on applicant reactions. There are also specific forms of looping (forced vs. choice looping and saves) that, while potentially important, were considered secondary to the broader feature of looping, thus mitigating their potential effect. For these reasons, convergence and specific forms of looping were not part of the theoretical models.

Unfortunately, one question that could not be answered via Study 1 was how applicant reactions are measured uniquely for branching (i.e., RQ4). Although some global reactions were measured regarding reactions to BSJTs, there was no measurement of applicant reactions regarding branching specifically. In this way, the current state of measurement of applicant reactions is like that described by Richman-Hirsch et al. (2000); although assessments that contain branching are viewed positively, it is difficult to parse reactions to branching from reactions to other test characteristics. Thus, newly developed reactions to branching measures or careful experimental control of branching characteristics is necessary to parse these effects. In Study 2, the latter approach will be utilized to quantitatively support the proposed models.

CHAPTER 4
STUDY 2

BSJTs are a relatively recent innovation in selection and assessment, and their popularity continues to grow despite a paucity of empirical literature supporting their use. The research that does exist is promising. Olson-Buchanan et al. (1998) appears to be the first study of a BSJT. The authors created and validated a BSJT intended to measure conflict resolution skills. Utilizing six different samples across a variety of industries, the researchers found that showed that the BSJT predicted manager ratings of on-the-job conflict resolution and ratings of overall performance while demonstrating discriminant validity with, and incremental validity above and beyond, tests of verbal skills and quantitative skills. Richman-Hirsh et al. (2000), using the same assessment developed by Olson-Buchannan et al., examined applicant reactions to three different assessment media: paper and pencil, text-based delivered via a computer, and a multi-media computer version. The authors found that the multimedia version was rated higher than the other versions in terms of content validity, predictive validity, job relevance, enjoyment, length, and satisfaction with the process. Although reactions to branching were not distinguished from reactions to the medium, this study did show that an SJT than contained branching could have a positive relationship with a number of important outcomes. Kanning et al. (2006) conducted two studies comparing reactions between SJTs and BSJTs, manipulating the medium of the scenario, the response options, and interactivity (i.e., SJT vs. BSJT). The authors found that the BSJT items were rated more positively compared to SJT items in terms of usefulness, emotional reaction, acceptance, fairness, and job relatedness. Thus, existing evidence indicates that BSJTs can be a valid assessment method that has a positive impact on applicant reactions.

Although helpful in terms of providing preliminary support for the use of BSJTs, existing studies on BSJTs used relatively simple forms of branching. As Study 1 demonstrated, branching can take many different forms. Specifically, there are four main design features of branching: contingency, parallelism, convergence, and looping. BSJTs can be driven solely by narrative progression, or branching can be contingent on test-taker performance. The different branches of a BSJT can be created so that the

constructs measured and the scenarios encountered are parallel and virtually the same, save for some strategically altered wording, or the content between branches can be dramatically different. Branches can converge through parent items which acts as a way to bring all test-takers back to a common item, or test-takers can be forced to certain items to ensure that they encounter certain, vital content. Finally, test-takers can be allowed to loop back to previous content, or they may only be able to follow a linear path through the test. Each of these different forms of branching could have a different impact on test-taker reactions.

Hausknecht et al. (2004) developed a comprehensive models of applicant reactions, detailing the relationships among perceived procedure characteristics (e.g., face validity, consistency administration, and opportunity to perform) and applicant perceptions such as motivation, perceived procedural justice, and attitudes toward the test. It was this model that served as the foundation for the theory developed in Study 1 and tested in Study 2. More specifically, it is theorized that the impact of various branching features on applicant perceptions occurs through perceived procedure characteristics. That is, perceived procedure characteristics mediate the relationship between branching features and application perceptions. The hypotheses detailed in the remainder of this section are all based on this theoretical model. All BSJTs are used to tell a story that follows a narrative that adapts to test-taker responses in some way. When a test-taker responds to a BSJT scenario, the next scenario furthers the story by presenting a scenario that is a realistic outcome of the choice the test-taker made to the first scenario (Olson-Buchanan et al., 1998). This story continues as the test-taker works through the test. Regardless of other branching features a BSJT utilizes, this story-like structure is likely to be present.

The intention of story-like BSJTs is to increase the likelihood that test-takers will feel it is realistic and face valid, to a greater degree than SJTs lacking branching. Indeed, on the continuum of presentation fidelity, branching is considered on the higher end of realism, just below live action (Tuzinski, 2013). Past research has shown that media richness has a positive impact on perceptions of face validity (Bruk-Lee, Drew, & Hawkes, 2013). Chan and Schmitt (1997), Motowidlo et al. (1990), and

Richman-Hirsh et al. (2000) all found that assessment media has a positive impact on perceptions of validity. Further, branching should also impact more distal outcomes such as motivation. A common theme that emerged from Study 1 was that the story-telling of a BSJT is intended to make candidates feel more immersed and engaged with the test. As the story unfolds, the test-taker's interest in the story should increase, as would their perception of their ability to impact the direction of the story.

This narrativization is also theorized to increase test-taker motivation. In addition to SME assertions, Roch et al. (2014) found that the use of semi-structured interviews, a form of branching assessment, in comparison to complete structure was positively related to motivation. Further, the interactivity of BSJTs, the responsiveness of the test to test-taker, is intended to give test-takers the sense that the content of the test is similar to content of the job. Test-takers should see the test as fairer compared to a traditional SJT because it more closely resembles the job. Thus, it is theorized that branching in general will lead to enhanced perceptions of procedural justice. While general branching is likely to directly impact perceptions of face validity and motivation, prior literature has shown that there are strong correlations among these outcomes. Hausknecht et al. (2004) found that face validity is strongly related to test motivation ($\rho = .35$) and procedural fairness ($\rho = .58$). Given these strong relationships, it is likely that the impact of branching on motivation and procedural justice is mediated by face validity. That is, to the extent that a test-taker feels a BSJT is more face valid than a traditional SJT, the BJST has a more positive impact on both motivation and perceptions of procedural justice.

Hypothesis 1. Face validity mediates the relationship between general branching and motivation such that general branching leads to greater face validity, which in turn leads to greater motivation.

Hypothesis 2. Face validity mediates the relationship between general branching and procedural justice such that general branching leads to greater face validity, which in turn leads to greater procedural justice.

In addition to branching in general impacting perceived procedure characteristics and applicant reactions, the specific types of branching (contingency, parallelism, and looping) should impact those

same outcomes. The contingency upon which branching occurs, either based on narrative flow or narrative flow and performance, has important implications. Branching based on test-taker performance is very similar to the mechanics of a CAT – as the test-taker answers items correctly, the test gets harder, and as the test-taker answers items incorrectly, the test gets easier. Research on the impact of CAT on test-taker reactions is mixed. Some authors have found that CATs have a positive impact on test-taker reactions (Sydell, Ferrell, Carpenter, Frost, & Brodbeck, 2013), and that using such tests to create a storyline experience can make the test seem more realistic (Parshall, Harmes, Davey, & Pashley, 2010). However, other research has shown that CAT can lead to more negative reactions. Tonidandel et al. (2002) found that test-takers of high ability had negative reactions to a CAT test due to the difficulty level of the questions, despite their ability level. For adaptive tests such as CATs or BSJTs, consistency of administration is likely to be an important perceived procedure characteristic. Contingency branching is driven by test-taker performance, and the content of the test is determined by how well or poorly the test-taker is doing. In such a circumstance, test-takers are likely to attend to the fact that the test is different among test-takers. Specifically, performance branching in addition to narrative is likely to cause test-takers to view the assessment as less consistent among test-takers. Further, a test that is seen as inconsistent among test-takers will be viewed negatively and as unfair.

Much like a CAT test, a BSJT is an assessment that by its nature has alternate forms. The degree to which these forms are similar is greatly impacted by the degree to which the content of the SJT items differ. A BSJT with parallel branches keeps the differences in content between different branches to a minimum, perhaps altering just a few words between branches. An example of this might be changing the job title of an individual referenced in a scenario to make it consistent with the response to the previous scenario. A BSJT with non-parallel branches is one where the content of different branches can vary greatly, changing characters, language, or context. Thus, two individuals taking a BSJT with parallel branching will experience two very similar tests, whereas two individuals taking a BSJT with non-parallel branching could have very different testing experiences. The consistency with which the test is

administered is a salient test characteristic. Although SJTs are considered to be very consistently administered (Bauer & Truxillo, 2006), BSJT are likely less so, particularly for non-parallel BSJTs.

Consistency of administration predicts a number of important applicant perceptions (Hausknecht et al., 2004). Dineen, Noe, and Wang (2004) found of five procedural justice characteristics, consistency of administration most strongly predicted perceived fairness. Ryan and Ployhart (1998) found that inconsistency in the assessment process, regardless of whether that inconsistency favors the test-taker or the organization, results in negative perceptions of fairness. Hausknect et al. (2004) found that consistency of administration is strongly related to both procedural justice ($\rho = .24$) and test attitude ($\rho = .50$). Given all of the above, it is theorized that consistency of administration will mediate the relationship between a) contingency branching and procedural justice, b) contingency branching and test attitude, c) parallel branching and procedural justice, and d) parallel branching and test attitude.

Hypothesis 3. Consistency of administration mediates the relationship between contingency branching and procedural justice such that narrative only branching leads to greater consistency of administration, which in turn leads to greater procedural justice.

Hypothesis 4. Consistency of administration mediates the relationship between contingency branching and test attitude such that narrative only branching leads to greater consistency of administration, which in turn leads to greater test attitude

.

Hypothesis 5. Consistency of administration mediates the relationship between parallel branching and procedural justice such that parallel branching leads to greater consistency of administration, which in turn leads to greater procedural justice.

Hypothesis 6. Consistency of administration mediates the relationship between parallel branching and test attitude such that parallel branching leads to greater consistency of administration, which in turn leads to greater test attitude.

Looping BSJTs allow test-takers to go back in the assessment and change their responses. This looping allows test-takers to explore the impact of choosing other responses. Additionally, looping may allow a test-taker to explore paths that are only tangential to the main assessment. That is, the paths do not contain scored scenarios but are still relevant to the choices that the test-taker makes. These alternate paths then loop back to the scored content. According to the interviews conducted for Study 1, being able to revisit previously completed portions should lead to a greater sense of opportunity to perform by giving test-takers the ability to explore alternate paths through the test and perform behaviors that are job relevant, and to learn more about the impact of their choices, and thus test-takers are likely to feel that they are able to truly show their knowledge and skills. The ability to loop is likely to give test-takers an enhanced sense of procedural justice through the ability to loop back in the test and explore the impact of alternate choices. A great deal of literature has shown that opportunity to perform is related to important applicant perceptions, including particularly high correlations with procedural justice (e.g., Hausknecht et al., 2004; McFarland, 2003; Schleicher et al., 2006). Given that looping is likely to have a positive impact on opportunity to perform, and that opportunity to perform in strongly correlated with procedural justice, it is likely that looping will positively impact perceptions of procedural justice through opportunity to perform.

Hypothesis 7. Opportunity to perform mediates the relationship between looping and procedural justice such that looping leads to greater opportunity to perform, which in turn leads to greater procedural justice.

Method

Participants. To test study hypotheses, the mediation testing software PROCESS was used. Given this, Preacher, Rucker and Hayes' (2007) recommendations were used to determine the sample size required to reach the appropriate statistical power for the tested models. The authors examined via a Monte Carlo simulation the Type I error rates and necessary sample sizes under various conditions to provide guidance regarding adequate sample for several different moderated mediation models. Because

the PROCESS analysis looks at all relationships simultaneously, power is determined by the smallest relationship in the model, which is then matched to the nearest more conservative estimate in the tables provided by the authors. Regarding the relationships among general branching, face validity, and motivation, the smallest theorized relationship identified was the one between branching and face validity (r = .23, Siskind, 2012). The same was true for the relationship among general branching, face validity, and procedural justice. Regarding the relationships among the specific branching features, the proposed moderators (consistency of administration and opportunity to perform), and the proposed outcomes (procedural justice and test attitude), the smallest relationship was between consistency of administration and test attitude (r = .21, Hausknecht et al., 2004). By interpreting Preacher et al.'s summary table for Model 4, a sample of 500 was therefore determined to be sufficient to detect the relationship between branching in general and face validity. The remaining proposed relationships were all estimated to be somewhat higher, including the correlations between face validity and test motivation (r = .31, Hausknecht et al.), and face validity and procedural justice (r = .50, Hausknecht et al.). For the relationships regarding specific branching features, a sample of 500 was found to be sufficient to detect the relationship between consistency of administration and test attitude. In this case, the closest effect reported by Preacher et al. was the same as the previous power analysis, so the recommended sample size did not change. The remaining proposed relationships were all estimated to be somewhat higher, including the correlations between contingency and consistency of administration (r = .23, Tonidandel et al., 2002), parallelism and consistency of administration (r = .32, Tonidandel & Quinones, 2000), looping and opportunity to perform (r = .37, Konradt, Warszta, & Ellwart, 2013), consistency of administration and procedural justice (r = .34, Hausknecht et al.), and opportunity to perform and procedural justice (r = .48, Hausknecht et al.).

With this intended sample size in mind, participants were recruited via Amazon's Mechanical Turk (MTurk), an on-line crowdsourcing website that allows individuals to complete short, discrete tasks to earn a monetary reward. MTurk has several advantages when used for data collection. Researchers are

able to review the quality of the work submitted by subjects, and can reject work based on its quality, which provides an added incentive for subjects to put in a good faith effort when completing the work. Additionally, MTurk samples tend to be more diverse than the typical sample of college students (Behrend, Sharek, Meade & Wiebe, 2011), and are at least as reliable (Burhmester, Kwang, & Gosling, 2011), which increases the generalizability of any results. MTurk is an ideal sample for this study for several reasons. It includes working-age adults who are likely to have some familiarity with taking selection assessments. Given the diversity of MTurk samples, it seems less likely to be range restricted on the variables of interest in this study, or to have characteristics that covary with these variables, in comparison to other convenience samples (such as undergraduates). Both characteristics are an important consideration when choosing a sample, as they can negatively affect external validity (Landers & Behrend, 2015).

Data collection occurred in two phases. After the MTurk HIT was posted, 605 people responded within 24 hours. After reviewing the data, 12 HITs were rejected based on missing all careless responding items. The remaining HITs were accepted and participants were paid $1.00 for their time. After a cursory review of data, it appeared that many cases would be lost after data cleaning. Because of this, a second HIT was posted. A total of 553 people responded to the second HIT, of which four people were rejected based on missing all careless responding items. Because participants were completing the HIT faster than anticipated, payment was reduced to $0.50 for this second round of data collection. To examine whether there were meaningful differences in the two samples, a MANOVA was run with sample membership as the independent variable and the six outcome measures (e.g., face validity, consistency of administration, etc.) as the dependent variables. There were no significant effects in either the multivariate omnibus test or in tests of individual effects, and the individual effect sizes were regardless all very small (all partial $\eta^2 < .003$). Thus, there were no differences between the two rounds of data collection regarding the study outcomes, and data collection round was not considered in any later analyses.

Overall, there were 1,174 submissions of the MTurk HIT. Of the 1,174 submissions, 19 were rejected (i.e., they were not compensated and their data was deleted) due to missing all three of the careless responding items (i.e., responding essentially at random). 49 were duplicate submissions. This left 1,106 submissions for further analysis. To assess careless responding, three bogus items were included per the recommendations of Meade and Craig (2012). Of remaining cases, 1043 participants missed no items, 41 participants missed 1 item, and 22 participants missed 2 items. To balance data integrity with the likelihood of removing careful responders, the 22 participants who missed 2 items were eliminated. Finally, scores on the manipulation check were examined. Of the remaining 1084 responders, 274 answered three items correctly, 392 answered two items correctly, 297 answered one item correctly, and 121 answered no items correctly. Given the subtlety of the manipulation and the relatively flat uniform distribution discovered here, a fairly stringent standard was imposed regarding the manipulation check. Specifically, participants were removed if they answered none or one item correctly. This resulted in 418 additional cases removed from the analysis, leaving a final sample size of 666 responders, representing 56.7% of the originally collected cases. The implications and potential causes of this will be discussed in detail in the discussion.

The final sample consisted of 666 respondents, who ranged in age from 19 to over 65 (mean age = 37). The sample was 52.9% Female (N = 352), 46.1% Male (N = 307), 0.6% Transgender (N = 4), and 0.5% other (N = 3). The sample was 79% European American or White (N = 529), 8.1% Asian American (N = 54), 5.7% African American or Black (N = 38), 3.5% Two or More Races (N = 23), 2.4% other (N = 16), 0.5% Pacific Islander or Native American (N = 3), 0.3% Native American or Native Alaskan (N = 2), and 0.2% Arab American. Sixty four percent of the sample was employed full time (N = 423), 18.9% was not employed (N = 126), and 16.2% was employed part-time (N = 108). For those employed full or part time, the average hours worked per week was 37.5. For those employed full or part time, the respondents were employed in a variety of industries, including 19.1% Other (N = 127), 12.2% in retail (N = 81), 11.3% in business services (N = 75), 8.7% in education (N = 58), 8% in healthcare (N = 53), 7.4% in

crowdsourcing (*N* = 49), 6.8% manufacturing (*N* = 45), 4.8% in finance (*N* = 32), 2.1% in insurance (*N* = 14), and 5% in wholesale. For those employed full or part time, 30.2% of the sample had been with their current organization for longer than five years (*N* = 201), 21.9% between three and five years (*N* = 146), 15.8% between 1 and years (*N* = 105), 6.9% between six months and one year (*N* = 46), and 6.2% for less than six months (*N* = 41). The majority of the sample were not MTurk Masters, with 13.2% having a masters designation (*N* = 88), and 86.6% not having a masters designation (*N* = 577). Most of the sample, 70.3%, completed MTurk HITs to supplement their primary income (*N* = 468), with 14.6% indicating that MTurk was their primary source of income (*N* = 97), 9.8% completing HITs for fun (*N* = 35), and 5.3% listing their reason as "other." On average, the sample spent an average of 14.7 hours per week completing MTurk HITs.

The PROCESS macro (Hayes, 2013) was used to evaluate hypotheses 1 through 7. PROCESS implements an OLS regression based path analytic framework to estimate both direct and indirect effects in mediational models. To enable analysis, Hayes provides a catalog of model templates that can be used to represent the researcher's models. In the present study, Hayes' model 4 was used to test seven empirical models, a prototype of which is depicted in the example Figure 4. Similar mediational models were tested for all study hypotheses.

Materials and Measures. Given the results of Study 1, the following stimulus materials and scales were identified to represent the constructs in the two core theoretical models (see Figures 2 and 3).

Descriptions of BSJTs. A series of bullet points representing types of branching were created based upon the theory developed in Study 1. A base test description was also written. To create the combinations of conditions derived from the study model, the base description was added to various bullet points. For example, for the control SJT condition, only the base description was used, but for the looping, non-parallel, narrative plus performance branching SJT, the base description plus the matching three bullets for those conditions were combined. These descriptions thus describe the nature of the

assessment, the way the assessment branches, and the intended purpose of the BSJT (i.e., for selection). These test descriptions are in Appendix C.

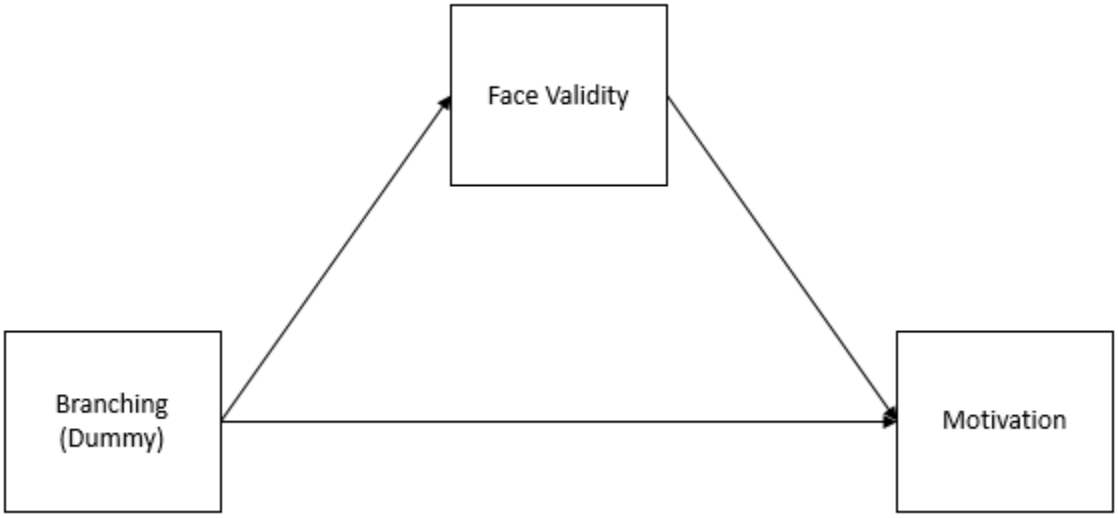

Figure 4. Theoretical model of the impact of branching on test motivation.

Perceived Procedure Characteristic: Face Validity. Face validity was measured using five items (α = .90) adapted from Smither, Reilly, Millsap, Pearlman, and Stoffey (1993). The authors reported α = .82 for their job relatedness scale, comprised of a predictive validity scale and a face validity scale, averaged across 14 measurement occasions. While no reliability estimate was provided for the face validity scale individually, Whitman, Kraus, and Van Rooy (2014) reported a reliability of α = .79. Example items include "I could not see any relationship between the test and what would be required on the job.", and "The actual content of the test would be clearly related to the job." All items on this scale, and all remaining scales to be described in this section, can be found in Appendix D.

Perceived Procedure Characteristic: Consistency of Administration. Consistency of administration was measured using three items (α = .85) from Bauer et al.'s (2001) Selection Procedural Justice Scale (SPJS). Example items include "The test would be administered to all applicants in the same way.", and "There would be no differences in the way the test was administered to different applicants."

Perceived Procedure Characteristic: Opportunity to Perform. Opportunity to perform was measured using four items (α = .95) from Bauer et al.'s (2001) SPJS. Example items include "I could really show my skills and abilities through this test.", and "This test would allow me to show what my job skills are."

Applicant Perception: Test Taking Motivation. Test taking motivation was measured using Arvey, Strickland, Drauden, & Martin's (1990) 10-item (α = .93) motivation scale. Example items include "I would try my best on this test.", and "I would be extremely motivated to do well on this test."

Applicant Perception: Attitude toward the Test. Attitude toward the test was measured using Tonidandel and Quinone's (2000) two item scale of test attitude (α =.94). Example items are "I would be in favor of having to take this type of test" and "I would not like to take this type of test."

Applicant Perception: Procedural Justice. Procedural Justice was measured using Colquitt's (2001) seven-item procedural (α =.82) justice scale. Example items include "Would those procedures be applied consistently?", and "Would those procedures be free of bias?

Manipulation Check. Three multiple choice questions were used to assess if each participant understood the condition to which they were assigned. Each item contained only one correct answer for each condition. As an example, for the parallel/nonparallel conditions, the item text was, "Based on the test description, choose the statement that would be true IF YOU WORKED THROUGH THE ACTUAL TEST: a) the test would present the same items in the same order to all candidates, b) there would be very little difference between the items you would see and the items other candidates would see, c) there could be large differences in the test content you see and the test content other candidates see, d) you could pick more than one response for each question. For the control condition, option A was correct. For the parallel

condition, option B was correct. For the nonparallel condition, option C was correct. Similarly, for the contingency item, option A was correct for the control condition, option B was correct for the narrative condition, and response option D was correct for the narrative plus performance condition. For the looping item, option D was correct for both the control and linear condition, and option B was correct for the looping condition. All items for this scale are located in Appendix E.

***Careless Responding*.** To detect careless responding, one item directing the respondent to answer an item a specific way was added to the face validity, motivation, and the procedural justice scale. The item for the face validity scale was "Choose the 4th response option (Somewhat like me) for this item." The item in the motivation scale was "Choose response option 7 (Strongly agree) for this item." The item in the procedural justice scale was "Choose response option 3 (the middle option) for this item."

Procedure. The MTurk HIT description informed subjects that the HIT was a research study examining test design for pre-employment testing. Subjects who accept the MTurk HIT were provided a notification statement. Then, similar to Kanning et al. (2006), each participant was randomly assigned to see one test description, randomly selected among the standard SJT or one of the eight descriptions of a BSJT. After the description, the relevant reactions measure were administered.

Results

The data were screened to ensure all variables are normal and lack outlying cases. Multivariate normality and outlyingness were assessed. To test the multivariate normality assumption, linearity and multicollinearity were examined per the recommendations of Kline (2005). To assess linearity, the relationship between every specified relationship was tested for a curvilinear relationship (i.e., quadratic, cubic, etc.) in SPSS using the regression curve estimation function. The F value for the linear relationship was significant and greater than the other relationships for all analyses, providing evidence that the relationships were sufficiently linear. To test for univariate outliers, the standardized values for all variables were calculated. Any values greater than plus or minus three was deemed as outlying. There were 12 outlying cases for the Motivation scale. Given that these cases only make up 2.1% of the sample,

they were left intact. To check for multivariate outlyingness, a series of linear regressions capturing Mahalanobis distances was run and significance was determined via a chi-square distribution. 4 cases were identified as statistically significant at the .001 level. Again, given their relative proportion to the entire sample (0.6%), these cases were left intact. The motivation scale was found to be negatively skewed, and so a log 10 transformation was performed.

To examine manipulation strength, a series of analyses were run to examine if responses to the three manipulation check items were related to branching features. Specifically, there was a "correct" answer to all three questions for participants in the control condition that implied no branching. Additionally, there was a single "correct" answer for each of the main effects (i.e., the first question was a manipulation check of contingency, the second of parallelism, and the third of looping). To test these effects, two sets of analyses were conducted. The results for all analyses are located in Table 1. First, to test the manipulation strength of the control condition, three independent samples t-tests were conducted to examine the effect of condition on correct responses to each manipulation check item. In all manipulation check analyses, Levene's test for homogeneity of variances was statistically significant, and thus the degrees of freedom for each t-test were adjusted to correct for heterogeneity. For item 1, the control condition scored significantly higher than the other conditions, $t(52.36) = 23.41$, $p < .001$. For item 2, the control condition scored significantly higher than the other conditions, $t(50.40) = 9.39$, $p < .001$. For item 3, the control condition scored significantly higher than the other conditions, $t(55.05) = 11.55$, $p < .001$. Second, to test the manipulation strength for the specific branching features, 6 independent samples t-tests were conducted, two for each manipulation check item. Specifically, the proportion of responses to each "correct" answer was examined comparing the condition for which it was the correct response with all other conditions. All tests were statistically significant and in the direction suggested by the manipulation check design. For item 1, the narrative condition scored significantly higher than the other conditions, $t(325.57) = 10.65$, $p < .001$. For item 1, the narrative plus performance condition scored significantly higher than the other conditions, $t(361.12) = 12.20$, $p < .001$. For item 2,

the parallel condition scored significantly higher than the other conditions, $t(438.95) = 18.90$, $p < .001$. For item 2, the non-parallel condition scored significantly higher than the other conditions, $t(483.37) = 18.83$, $p < .001$. For item 3, the looping condition scored significantly higher than the other conditions, $t(382.09) = 47.83$, $p < .001$. For item 3, the linear condition scored significantly higher than the other conditions, $t(578.12) = 52.43$, $p < .001$. Taken together, these results indicate that the manipulation was experienced by participants, but variance in effect sizes suggests it varied widely in its effect.

Table 1
Means, Standard Deviations, *t* Values and Effect Sizes for Manipulation Check Items.

Condition	Manipulation Check Item	N_1	M_1	SD_1	N_2	M_2	SD_2	*t*	*d*
Control vs Branching	1	51	.92	.27	615	.02	.14	23.41	5.73
Control vs Branching	2	51	.65	.48	615	.01	.11	9.39	3.80
Control vs Branching*	3	51	.76	.43	292	.05	.23	11.55	2.66
Narrative	1	270	.36	.48	345	.03	.18	10.65	.95
Narrative+Performance	1	345	.94	.23	270	.54	.50	12.20	1.07
Parallel	2	275	.69	.46	340	.09	.29	18.90	1.60
Non-parallel	2	340	.88	.32	275	.28	.45	18.83	1.58
Looping	3	292	.90	.29	323	.02	.12	47.83	4.00
Linear	3	323	.96	.20	292	.05	.23	52.43	4.27

Note. All tests were statistically significant with $p < .001$. Group 1 = participants in the condition targeted with that manipulation check item; Group 2 = all other participants. *This test removed linear condition as the correct response was the same as the control condition.

Although it was not hypothesized, a series of ANOVAs were run to detect any two or three way interactions for all study relationships, because the presence of such interactions would confound the interpretation of main effects. None of these ANOVAs indicated that there were any such interactions, so the remainder of analyses focused upon hypothesized direct and indirect effects only.

To test focal hypotheses, Hayes' (2013) PROCESS macro was used to assess direct and indirect effects. Each model was tested separately, and statistically significant path coefficients indicated statistically significant causal direct effects when a manipulation was the IV. The PROCESS macro multiplied direct path coefficients within mediational pathways and computed bias-corrected bootstrapped

confidence intervals around those indirect effect estimates to determine statistical significance. Specifically, when confidence intervals included zero, they were interpreted to indicate statistically significant indirect effects. All direct and indirect effects depicted in Figure 3 were tested with their matched PROCESS procedure.

Table 2
Means, Standard Deviations, and Correlations for Predictors and Outcomes.

Scale	Mean	SD	1	2	3	4	5	6	7	8	9	10
1. Branching	.85	.53	-									
2. Contingency	-.12	.99	-	-								
3. Parallel	-.11	1.00	-	-.09*	-							
4. Looping	-.05	1.00	-	.04	-.05	-						
5. Face Validity	3.66	0.89	-.04	.08	-.02	.04	(.90)					
6. CoA	3.36	1.02	-.20**	.05	.14	.02	.14**	(.85)				
7. OtP	3.37	0.98	.08	.08	.00	.06	.45**	.06	(.95)			
8. Proc. Justice	3.08	0.77	.00	.04	.08	-.07	.36**	.17**	.59**	(.82)		
9. Test Attitude	3.20	1.15	.04	.07	.00	-.01	.49**	.09*	.65**	.60**	(.94)	
10. Motivation	5.58	.84	.01	.05	-.01	.03	.40**	.16**	.37**	.26**	.30**	(.93)

Note. * p < .05; ** p < .01; N = 613 – 666; CoA = consistency of administration, OtP = opportunity to perform

To test Hypotheses 1 - 7, a series of PROCESS analyses were conducted. Hypothesis 1 stated that face validity would mediate the relationship between general branching and motivation. The model results are detailed in Table 3, and results for the test of indirect effects for all hypotheses appear in Table 4. Results indicated that general branching was not significantly related to face validity (b = -0.064, p = .326), face validity was significantly negatively related to motivation (b = -0.068, p < .001), general branching was not significantly related to motivation after controlling for face validity (b = -0.001, p = .329), and face validity did not mediate the relationship between general branching and motivation, b = -0.022, SE = 0.023 CI [-0.066, 0.025], failing to support Hypothesis 1.

Table 3
PROCESS Analysis for Control/Branching, Face Validity, and Motivation.

		Outcome						
		Face Validity				Motivation		
Predictor		Coeff.	*SE*	*p*		Coeff.	*SE*	*p*
Condition	*a*	-0.064	0.065	.326	*c'*	-0.01	0.104	.329
Face Validity		-	-	-	*b*	-0.068	0.006	< .001
Constant	i_1	3.709	0.065	< .001	i_2	0.524	0.025	< .001
		R^2 = .002				R^2 = .160		
		F(1,664) = 0.966, *p* = .326				F(1,660) = 61.198, *p* < .001		

Note. Control coded as -1, Branching coded as 1.

Table 4
Summary of Indirect Effects.

Branching Feature	Mediator	Outcome	*b*	*SE*	LLCI	ULCI
General Branching	Face validity	Motivation	-0.022	0.023	-0.066	0.025
General Branching	Face validity	Procedural Justice	-0.020	0.021	- 0.058	0.018
Contingency	Consistency of Administration	Procedural Justice	0.007	0.006	-0.005	0.020
Contingency	Consistency of Administration	Test Attitude	0.006	0.006	-0.005	0.020
Parallelism	Consistency of Administration	Procedural Justice	0.020	0.008	0.007	0.037
Parallelism	Consistency of Administration	Test Attitude	0.019	0.009	0.004	0.040
Looping	Opportunity to Perform	Procedural Justice	0.026	0.019	-0.010	0.062

Hypothesis 2 stated that face validity would mediate the relationship between general branching and procedural justice. Model results are detailed in Table 5. Results indicated that general branching was not significantly related to face validity (b = -0.064, p = .326), face validity was significantly positively related to procedural justice (b = 0.318, p < .001), general branching was not significantly related to procedural justice after controlling for face validity (b = 0.025, p = .638), and face validity did not mediate the relationship between general branching and procedural justice, b = -0.02, SE = 0.019 CI [-0.058, 0.018], failing to support Hypothesis 2.

Hypothesis 3 stated that consistency of administration would mediate the relationship between contingency branching and procedural justice. Model results are detailed in Table 6. Results indicated that contingency branching was not significantly related to consistency of administration (b = 0.047, p = .260), consistency of administration was significantly positively related to procedural justice (b = 0.141, p <

.001), contingency branching was not significantly related to procedural justice after controlling for consistency of administration (b = 0.022, p = .491), and consistency of administration did not mediate the relationship between contingency branching and procedural justice, b = 0.007, SE = 0.006 CI [-0.051, 0.02], failing to support Hypothesis 3.

Table 5
PROCESS Analysis for Control/Branching, Face Validity, and Procedural Justice.

		Outcome						
		Face Validity				Motivation		
Predictor		Coeff.	*SE*	*p*		Coeff.	*SE*	*p*
Condition	*a*	-0.064	0.065	.326	*c'*	0.025	0.053	.638
Face Validity		-	-	-	*b*	0.318	0.032	< .001
Constant	i_1	3.709	0.065	< .001	i_2	1.894	0.128	< .001
		R^2 = .002				R^2 = .133		
		F(1, 664) = 0.966, p = .326				F(2,663) =50.67, p < .001		

Note. Control coded as -1, Branching coded as 1.

Table 6
PROCESS Analysis for Contingency, Consistency of Administration, and Procedural Justice.

		Outcome						
		Consistency of Administration				Procedural Justice		
Predictor		Coeff.	*SE*	*p*		Coeff.	*SE*	*p*
Contingency	*a*	0.047	0.042	.260	*c'*	0.022	0.031	.491
Consistency of Administration		-	-	-	*b*	0.141	0.03	< .001
Constant	i_1	3.309	0.042	< .001	i_2	2.613	0.105	< .001
		R^2 = .002				R^2 = .036		
		F(1,613) = 1.271, p = .26				F(2,612) = 11.287, p < .001		

Note. Narrative coded as -1, Narrative plus Performance coded as 1.

Hypothesis 4 stated that consistency of administration would mediate the relationship between contingency branching and test attitude. Model results are detailed in Table 7. Results indicated that contingency branching was not significantly related to consistency of administration (b = 0.044, p = .294), consistency of administration was significantly positively related to test attitude (b = 0.127, p = .005),

contingency branching was not significantly related to test attitude after controlling for consistency of administration (b = 0.071, p = .127), and consistency of administration did not mediate the relationship between contingency branching and test attitude, b = 0.006, SE = 0.006 CI [-0.005, 0.020], failing to support Hypothesis 4.

Table 7
PROCESS Analysis for Contingency, Consistency of Administration, and Test Attitude.

		Outcome						
		Consistency of Administration				Test Attitude		
Predictor		Coeff.	*SE*	*p*		Coeff.	*SE*	*p*
Contingency	*a*	0.044	0.042	.294	*c'*	0.071	0.046	.127
Consistency of Administration		-	-	-	*b*	0.127	0.045	.005
Constant	i_1	3.312	0.042	< .001	i_2	2.801	0.156	< .001
		R^2 = .002				R^2 = .017		
		F(1,612) = 1.105, p = .294				F(2,611) = 5.312, p = .005		

Note. Narrative coded as -1, Narrative plus Performance coded as 1.

Hypothesis 5 stated that consistency of administration would mediate the relationship between parallel branching and procedural justice. Model results are detailed in Table 8. Results indicated that parallel branching was significantly related to consistency of administration (b = 0.047, p < .001), such that parallel branching was seen as more consistent than non-parallel branching, consistency of administration was significantly positively related to procedural justice (b = 0.136, p < .001), parallel branching was not significantly related to procedural justice after controlling for consistency of administration (b = 0.044, p = .162), and consistency of administration did mediate the relationship between parallel branching and procedural justice, supporting Hypothesis 5, b = 0.020, SE = 0.008 CI [0.007, 0.037], such that parallel branching led to increased consistency of administration, which in turn led to increased procedural justice.

Table 8
PROCESS Analysis for Parallel Branching, Consistency of Administration and Procedural Justice.

		Outcome						
		Consistency of Administration				Procedural Justice		
Predictor		*Coeff.*	*SE*	*p*		*Coeff.*	*SE*	*p*
Parallel Branching	*a*	0.147	0.041	< .001	*c'*	0.044	0.031	.162
Consistency of Administration		-	-	-	*b*	0.136	0.031	< .001
Constant	i_1	3.319	0.041	< .001	i_2	2.632	0.106	< .001
		R^2 = .020				R^2 = .038		
		$F(1,613) = 12.784$, $p < .001$				$F(2,612) = 12.055$, $p < .001$		

Note. Nonparallel coded as -1, Parallel coded as 1.

Hypothesis 6 stated that consistency of administration would mediate the relationship between parallel branching and test attitudes. Model results are detailed in Table 9. Results indicated that parallel branching was significantly related to consistency of administration ($b = 0.144$, $p < .001$) such that parallel branching was seen as more consistent than non-parallel branching, consistency of administration was significantly positively related to test attitude ($b = 0.132$, $p = .004$), parallel branching was not significantly related to procedural justice after controlling for consistency of administration ($b = -0.017$, $p = .711$), and consistency of administration mediated the relationship between parallel branching and test attitude, supporting Hypothesis 6, $b = 0.019$, SE = 0.009 CI [0.004, 0.04] such that parallel branching led to increased consistency of administration in comparison to non-parallel branching, which in turn led to increased test attitude.

Hypothesis 7 stated that opportunity to perform would mediate the relationship between looping and procedural justice. Model results are detailed in Table 10. Results indicated that looping was not significantly related to opportunity to perform ($b = 0.053$, $p = .176$), opportunity to perform was significantly positively related to procedural justice ($b = 0.48$, $p < .001$), looping was negatively significantly related to test procedural justice after controlling for opportunity to perform ($b = -0.076$, $p = .003$) such that looping led to a decreased sense of procedural justice, and opportunity to perform did not mediate the relationship between looping and procedural justice, $b = 0.026$, SE = 0.019 CI [-0.01, 0.062], failing to support Hypothesis 7.

Table 9
PROCESS Analysis for Parallel Branching, Consistency of Administration, and Test Attitude.

		Outcome						
		Consistency of Administration				Test Attitude		
Predictor		Coeff.	*SE*	*p*		Coeff.	*SE*	*p*
Parallel Branching	*a*	0.144	0.041	< .001	*c'*	-0.017	0.047	.711
Consistency of Administration		-	-	-	*B*	0.132	0.046	.004
Constant	i_1	3.322	0.041	< .001	i_2	2.773	0.158	< .001
		R^2 = .020				R^2 = .014		
		$F(1,612) = 12.289$, $p < .001$				$F(2,611) = 4.198$, $p = .016$		

Note. Nonparallel coded as -1, Parallel coded as 1.

Table 10
PROCESS Analysis for Looping, Opportunity to Perform, and Procedural Justice.

		Outcome						
		Opportunity to Perform				Procedural Justice		
Predictor		Coeff.	*SE*	*p*		Coeff.	*SE*	*p*
Looping	*a*	0.053	0.039	.176	*c'*	-0.076	0.025	.003
Opportunity to Perform		-	-	-	*b*	0.480	0.026	< .001
constant	i_1	3.400	0.039	< .001	i_2	1.443	0.092	< .001
		R^2 = .003				R^2 = .361		
		$F(1,613) = 1.832$, $p = .176$				$F(2,612) = 173.002$, $p < .001$		

Note. Linear coded as -1, Looping coded as 1.

Discussion

The primary goal of Study 2 was to empirically examine the impact of BSJTs on applicant reactions. Specifically, Study 2 examined the mediating effects of perceived procedure characteristics on the relationship between general and specific forms of branching and applicant perceptions. Regarding the impact of branching in general, results indicated that branching in general did not impact test taking motivation or procedural justice, either directly or through face validity. This finding directly contradicts the idea that branching impacts applicant reactions via its impact on face validity. However, as Table 2 shows, there was a significant negative correlation between general branching and consistency of administration (r = -.20). To further explore this relationship, an additional PROCESS analysis was run to

examine the mediating effect of consistency of administration on the relationship between general branching and motivation and between general branching and procedural justice. Model results are detailed in tables 11 and 12, and indirect effects are detailed in table 13. Results indicated that consistency of administration mediated both relationships: b = 0.009, SE = 0.003 CI [0.004, 0.015] for motivation, and b = -0.051, SE = 0.010 CI [-0.068, -0.028] for procedural justice. While general branching led to an overall decreased perception of consistency of administration, it led to increased test-taker motivation and a decreased sense of fairness.

Table 11
PROCESS Analysis for Control/Branching, Consistency of Administration, and Motivation.

		Outcome						
		Consistency of Administration				Motivation		
Predictor		Coeff.	*SE*	*p*		Coeff.	*SE*	*p*
Condition	*a*	-0.388	0.073	< .001	*c'*	-0.015	0.011	.177
Consistency of Administration		-	-	-	*b*	-0.024	0.006	< .001
Constant	i_1	3.692	0.073	< .001	i_2	0.358	0.025	< .001
		R^2 = .041				R^2 = .024		
		$F(1,661) = 27.921$, $p < .001$				$F(2,660) = 8.103$, $p < .001$		

Note. Control coded as -1, Branching coded as 1.

For specific forms of branching, contingency did not impact procedural justice or test attitudes, either directly or through consistency of administration. Consistency of administration did mediate the relationship between parallel branching and procedural justice and between parallel branching and test attitude. Specifically, BSJTs with parallel content were seen as more consistent, which in turn led to more positive perceptions of procedural justice. Finally, opportunity to perform did not mediate the relationship between looping and procedural justice.

Table 12
PROCESS Analysis for Control/Branching, Consistency of Administration, and Procedural Justice.

		Outcome						
		Consistency of Administration				Procedural Justice		
Predictor		Coeff.	*SE*	*p*		Coeff.	*SE*	*p*
Condition	*a*	-0.388	0.073	< .001	*c'*	0.056	0.057	.324
Consistency of Administration		-	-	-	*b*	0.133	0.03	< .001
Constant	i_1	3.691	0.073	< .001	i_2	2.582	0.122	< .001
			R^2 = .041				R^2 = .03	
			$F(1,664) = 28.262$, $p < .001$				$F(2,663) = 10.068$, $p < .001$	

Note. Control coded as -1, Branching coded as 1.

Table 13
Summary of Post Hoc Indirect Effects.

Branching Feature	Mediator	Outcome	*b*	*SE*	LLCI	ULCI
General Branching	Consistency of Administration	Motivation	0.009	0.003	0.004	0.015
General Branching	Consistency of Administration	Procedural Justice	-0.051	0.015	- 0.082	-0.024

Applicant reactions have an impact on important attitudes, intentions, and behaviors (McCarthy, Bauer, Truxillo, Anderson, Costa, & Ahmed, 2017). One of the main motivators behind the adoption of BSJTs is that the various branching innovations have a positive impact on applicant reactions. This study has shown that some of these innovations do positively impact application reactions. Specifically, parallel branching, where the differences between different branches are kept to a minimum, had a positive impact on consistency of administration in comparison to nonparallel branching, which in turn led to a parallel test being seen as fairer and also towards a more positive test attitude. Both applicant perceptions impact important organizational outcomes, including organizational attractiveness, recommendation intentions, offer acceptance intentions, and self-efficacy (Hausknecht et al., 2004).

Although looping was not related to procedural justice via opportunity to perform, it was directly related to procedural justice, *b* = -0.076, SE = 0.025 CI [-0.125 -0.026]. To examine the possibility that there was a different mediator, PROCESS analyses were run exploring the mediating effect of face validity and consistency of administration. Neither of these analyses were significant. This suggests that

there is another mediator impacting the relationship between looping and procedural justice. Reconsideration opportunity was one mediator that was considered in Study 1 but was not tested in Study 2 due to its relatively infrequent considering by SMEs. However, given that looping allows test-takers to go back in the test, this could be seen primarily as reconsideration by test-takers. Future research should explore the possibility that it is reconsideration opportunity rather than opportunity to perform that mediates the relationship between looping and procedural justice.

Limitations. The current study has several limitations. First, rather than completing actual BSJTs, respondents read descriptions of BSJTs. These descriptions included explicit acknowledgement of the ways the test branched and the impact of branching on test structure or scores. Realistically, test administrators might not acknowledge that an assessment branches or the potential impact of that branching. Further, some features, like parallel branching, might not even be noticed by test-takers unless explicitly noted by the test creator. Test creators may want to communicate to test-takers the kinds of branching they utilize to ensure they realize the positive impact on applicant reactions that the branching is intended to elicit. However, there are positives to the current study design that mitigate this limitation. Explicitly stating the test characteristics and their implications to the respondents provided a stronger manipulation than having test-takers complete an actual BSJT. If the current study did not find any significant results, given the strength of the manipulation, then it would be unlikely that effects would be found if a real BSJT were administered without the corresponding explanations. Thus, this study is perhaps better viewed as an important first step toward more subtle, higher fidelity research designs with research priorities derived from the current results.

Another limitation is the lack of a selection outcome. In a real-world setting, hiring decisions are based on candidate test scores. A great deal of research has shown that positive and negative decisions impact applicant reactions (e.g., Elkins & Phillips, 2000; Horvath, Ryan, & Stierwalt, 2000; Ryan, Sacco, McFarland, & Kriska, 2000). Thus, the findings in the current study might be different depending on whether someone was actually hired or not based on a BSJT. For example, a test-taker not hired after

taking BSJT that had parallel branching might find even the smallest differences to be a source of negative reactions. Further research should explore the impact of selection decision on the relationships examined in this study.

Finally, the t-tests for the manipulation checks generally showed that group targeted by the manipulation in each condition answered the items consistently with the intended effect. However, the effect sizes for the contingency groups (d = 0.95 for narrative and d = 1.07 for narrative plus performance) and for the parallel groups (d = 1.60 for parallel and d = 1.58 for non-parallel) were smaller than the effects of other conditions. Effect sizes, shown in Table 1, show that the groups sometimes answered in the non-targeted direction quite consistently; for example, although 94% of participants in the narrative plus performance condition answered in the direction of narrative plus performance, 54% of participants in the narrative condition also answered in this direction. Particularly problematic was the narrative condition, in which only 36% of participants in the narrative condition responded with the "narrative" answer (although only 3% of participants in the narrative plus performance conditions responded with that answer). This pattern could indicate that some responders had difficulty detecting the manipulation, that the manipulation check items did not validly measure manipulation strength, or that some combination of these two effects occurred. Ultimately, this makes the manipulation check difficult to interpret. Future research should make sure that the differences between these groups is clear, and that the manipulation check items accurately capture the important details of the relevant BSJT description.

CHAPTER 5

GENERAL DISCUSSION

Although SJTs have been in use for almost a century (McDaniel et al., 2001), BSJTs are a recent innovation. The goal of this study was to learn more about the different features of BSJTs and how these features impact important outcomes. The pilot study provided insight into BSJT usage, most importantly revealing that the primary motivation behind the increased usage of BSJTs is their perceived positive impact on applicant reactions. Study 1, using a grounded theory approach, explored this motivation more deeply through a series of interviews with BSJT SMEs. Through these interviews, several themes emerged regarding a) the different features of BSJTs, and b) the expected impact of these features on applicant reactions, resulting in a theoretical model of the impact of branching features on applicant reactions. Based on those interviews and empirical literature, it was theorized that general and specific branching features impact applicant perceptions via perceived procedure characteristics. Study 2 empirically tested the theoretical models.

The results of Study 2 showed that while some parts of the theoretical model were supported, others were not. While it was hypothesized that face validity would mediate the relationship between general branching and motivation, as well as the relationship between general branching and procedural justice, it was actually consistency of administration that mediated these relationships. Specifically, a difference in branching alone led to a decreased sense of test consistency, and decreased consistency perceptions were associated with increased test motivation but a decreased sense of fairness. It was hypothesized that consistency of administration would mediate the relationship between contingency branching and procedural justice and test attitude, but the data did not support this. The relationship between parallel branching and procedural justice, as well as the relationship between parallel branching and test attitude, were also mediated by consistency of administration. In both cases, parallel branching led to an increased sense of consistency compared to non-parallel branching, which led to a more positive sense of fairness and a more positive test attitude for those anticipating parallel branches.

Finally, looping had a positive impact on perceptions of fairness, though this relationship was not mediated by opportunity to perform as it was hypothesized.

These results have important practical implications. The finding that general branching had a positive effect on motivation but a negative effect on fairness indicated that, contrary to a large proportion of the SMEs interviewed for Study 1, the impact of branching in general may not be uniformly positive. While branching may lead to increased test-taker motivation, it can also lead to a decreased sense of fairness. From a test creator perspective, this means that branching shouldn't be implemented without some consideration of the desired outcome. If test creators are interested in increasing test-taker motivation, then branching may help achieve that goal. However, if test creators are concerned about creating a sense of fairness, then branching may in fact have the opposite effect.

According to Study 2, this negative reaction centered on the use of nonparallel BSJTs. This suggests that test creators should try to make the different branches as similar as possible, as parallel branching leads to increased perceptions of fairness and test attitude. However, given practical constraints, test creators may need to create branches that vary greatly in their content. In such cases, although test-takers may share some information about branching with test-takers to improve applicant reactions, they could avoid explicitly mentioning that test-takers might see very different versions of the test. From the present results, it appears that test creators should highlight the story-like nature of BSJTs, as this appears to have a positive impact on test-taker motivation.

That contingency branching did not have a relationship with applicant reactions may actually be a positive result from a test creator's perspective. Given that CATs can sometimes result in negative applicant reactions (e.g., Tonidandel et al., 2002), it could be that the story-like nature of a BSJT mitigates potential negative reactions. Thus, while contingency branching may not positively impact applicant reactions, if it keeps reactions from being negative despite an adaptive design, that is still a beneficial effect. As Bruk-Lee et al. (2013) pointed out, reactions to adaptive tests with a story-like structure are positive. Certainly there are some adaptive tests, like the Graduate Record Exam, where it

might not be possible to build a story into the assessment. However, with some planning and effort, it would certainly be possible to build a narrative into an adaptive test. For example, test vendors create huge item banks from which items are chosen for CATs. It may be possible to add a story-like element to the test or the items.

While the relationship between looping and procedural justice was not mediated by opportunity to perform, there was a significant negative relationship between the two variables. This finding has important implications for the construction of BSJTs. Broadly, looping may negatively impact applicant reactions while also increasing item exposure, shortening the life of the assessment. This was surprising, given that the practitioners who discussed this feature explicitly stated that they felt that allowing looping would result in more positive reactions. It could be that allowing test takers to go back and make different choice might make any mistakes more salient, and thus lead to decreased reactions. Absent from the current study was the specific looping feature of 'saving', which would result in a higher test score for a test taker. Such looping would be expected to lead to more positive reactions. Practitioners should consider carefully the kind of looping they employ and its potential impact on reactions before building it into a BSJT, and future research on looping is warranted.

In Study 1, a SME mentioned that for the tests that he had experience with, candidates who looped back could change their responses and receive partial credit. If looping does in fact lead to more positive perceptions of fairness, then allowing test-takers to change their answers to potentially boost their score might increase the magnitude of this effect. However, allowing test-takers to change responses could have an impact on other test characteristics (e.g., reliability, validity), so this innovation should be implemented with caution and requires further research before recommendations can be made confidently.

Study 2 revealed a number of significant relationships between branching features and applicant reactions. A close look at Tables 3-12 indicates that, while statistically significant, the magnitude of these relationships are not particularly large. In general, the largest mediating effects explain less than 2% of

the variance in application reactions. The exception is the 36% of the variance in procedural justice explained by looping and opportunity to perform. A close examination of the coefficients in Table 10 reveal that opportunity to perform has a particularly strong positive relationship with procedural justice (*b* = 0.48). However, BSJTs are still likely to be of value if they serve as a supplement or replacement for more expensive selection tests, such as assessment centers which often require a great deal of expense to create, administer, and score. Further, in a high volume testing environment, the impact of even a small improvement in applicant reactions can have a profound impact.

Future research can benefit from examining the impact of different branching features on other important test characteristics such as reliability, validity, and adverse impact. While applicant reactions are an important area of research, the value of BSJTs as a selection device is also tied to their ability to predict job performance reliably and in a way that doesn't disadvantage a particular group of test-takers. With regard to reliability, it is important for both researchers and practitioners to better understand the optimal way to assess reliability. SJT reliability is typically assessed via test-retest reliability, yet this may not be the best way to assess the reliability for BSJTs if different questions are seen by different test-takers and especially if test-takers can return to revise questions they believe they answered poorly. Additional research is needed to address this important issue.

Future research designs could also be improved to ensure that the differences between the features are clear. The results of the manipulation check analyses, especially combined with the relatively large number of cases eliminated for careless responding related to failing checks, indicated that the strength of the manipulations among all conditions may have been weak. Even among those participants that did not fail the careless responding screen, the manipulation strength of the contingency conditions and parallel conditions appeared to have been even weaker than that of the other conditions. Altogether, these weak manipulations might have made it difficult to for participants to remember or experience the intended differences between the conditions. One potential way to overcome this issue would be to use a within-subjects design, so that each participant sees the description for each branching feature rather than

just one. In such a design, the differences among the branching features might be more obvious, strengthening the manipulation. However, an alternative explanation is that even if practitioners describe specific branching features to job applicants, job applicants might also not attend to such information closely. Specifically, if practitioners intend to increase test motivation or reactions by describing the branching in an upcoming assessment, such descriptions may be difficult for job applicants to parse. Further research is thus required to understand the impact and interpretability of such descriptions before completing a branching assessment.

Overall, this study demonstrated that communicating with test-takers about the use of branching can, under some circumstances, impact anticipated reactions. Although this is not the first empirical study of BSJTs, it is the first to develop a framework for the many ways that SJTs can branch and test this framework empirically. Specifically, this study showed that the various branching design features impact outcomes in different ways. It is my hope that this study spurs an increase in the empirical interest in BSJTs, building upon the present results.

REFERENCES

Arthur, W., Jr., & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology, 93,* 435-442.

Arvey, R. D., Strickland, W., Drauden, G., & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology,* 43, 695-716.

Bauer, T. N., Maertz, C. P., Jr., Dolen, M. R., & Campion, M. A. (1998). Longitudinal assessment of applicant reactions to employment testing and test outcome feedback. *Journal of Applied Psychology, 83,* 892-903.

Bauer, T. N., & Truxillo, D. M. (2006). Applicant reactions to situational judgment tests. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 233–249). Mahwah, NJ: Erlbaum.

Bauer, T. N., Truxillo, D. M., Sanchez, R., Craig, J., Ferrara, P., & Campion, M. A. (2001). Development of the selection procedural justice scale (SPJS). *Personnel Psychology, 54,* 387-419.

Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods, 43,* 800-813.

Bledow, R., & Frese, M. (2009). A situational judgment test of personal initiative and its relationship to performance. *Personnel Psychology, 62,* 229-258.

Boyce, A. S., Corbet, C. E., & Adler, S. (2013). Simulations for personnel selection: An introduction. In M. Fetzer & K. Tuzinski (Eds.), *Simulations for personnel selection* (pp. 17-41).

Bruk-Lee, V., Drew, E. N., & Hawkes, B. (2013). Candidate reactions to simulations and media-rich assessments in personnel selection. In M. Fetzer & K. Tuzinski (Eds.), *Simulations for personnel selection* (pp. 43-60).

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6,* 3-5.

Chan, D. (1997). Racial subgroup differences in predictive validity perceptions on personality and cognitive ability tests. *Journal of Applied Psychology, 82,* 311-320.

Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology, 82,* 143-159.

Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Harvey, V. S. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology, 86,* 410-417.

Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology, 86,* 386-400.

Dineen, B. R., Noe, R. A., & Wang, C. (2004). Perceived fairness of web-based applicant screening procedures: Weighing the rules of justice and the role of individual differences. *Human Resource Management, 43,* 127-145.

Elkins, T. J., & Phillips, J. S. (2000). Job context, selection decision outcome, and the perceived fairness of selection tests: Biodata as an illustrative case. *Journal of Applied Psychology, 85,* 479-484.

Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review, 18,* 694-734.

Gilliland, S. W. (1994). Effects of procedural and distributive justice on reactions to a selection system. *Journal of Applied Psychology, 79,* 691-701.

Gilliland, S.W., Groth, M., Baker, R. C., Dew, A. F., Polly, L. M., & Langdon, J. C. (2001). Improving applicants' reactions to rejection letters: An application of fairness theory. *Personnel Psychology, 54,* 669−703.

Gilliland, S. W., & Hale, J. M. S. (2005). How can justice be used to improve employee selection practices? In J. Greenberg & J.A. Colquitt (Eds.), *Handbook of Organizational justice* (pp. 411−438). Mahwah, NJ: Lawrence Erlbaum.

Glaser, B. G. (2010). The future of grounded theory. *The Grounded theory Review: An International Journal, 9,* 1-14.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research.* New York, NY: Aldine.

Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology, 57,* 639-683.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis*. New York, NY: Guilford Press.

Horvath, M., Ryan, A. M., & Stierwalt, S. L. (2000). The influence of explanations for selection test use, outcome favorability, and self-efficacy on test-taker perceptions. *Organizational Behavior and Human Decision Processes, 83,* 310-330.

HumRRO. (n.d.) Update on Rich-Media Simulations. Retrieved from https://www.humrro.org/corpsite/press-release/update-rich-media-simulations.

Isabella, L. A. 1990. Evolving interpretations as change unfolds: How managers construe key organizational events. A*cademy of Management Journal, 33,* 7–41.

Kanning, U. P., Grewe, K., Hollenberg, S., & Hadouch, M. (2006). From the subjects' point of view: Reactions to different types of situational judgment items. *European Journal of Psychological Assessment, 22,* 168-176.

Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling*. New York, NY: The Guilford Press.

Konradt, U., Warszta, T., & Ellwart, T. (2013). Fairness perceptions in web-based selection: Impact on applicants' pursuit intentions, recommendation intentions, and intentions to reapply. *International Journal of Selection and Assessment, 21,* 155-169.

Kreiner, G. E., Hollensbe, E. C., & Sheep, M. L. (2009). Balancing borders and bridges: Negotiating the work-home interface via boundary work tactics. *Academy of Management Journal, 52,* 704–730.

Landers, R. N., & Behrend, T. S. (2015). An inconvenient truth: Arbitrary distinctions between organizational, mechanical turk, and other convenience samples. *Industrial and Organizational Psychology, 8*, 142-164.

Lievens, F., Buyse, T., & Sackett, P. R. (2005). The operational validity of a video-based situational judgment test for medical college admissions: Illustrating the importance of matching predictor and criterion construct domains. *Journal of Applied Psychology, 90*, 442-452.

Lievens, F., & Sackett, P. R. (2006). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology, 91*, 1181-1188.

Lievens, F., Sackett, P. R. & Buyse, T. (2009). The effects of response instructions on situational judgment test performance and validity in a high-stakes context. *Journal of Applied Psychology, 94*, 1095-1101.

Lievens, F., & Motowidlo, S. J. (2016). Situational judgment tests: From measures of situational judgment to measures of general domain knowledge. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 9*, 3-22.

Macan, T. H., Avedon, M. J., Paese, M., & Smith, D. E. (1994). The effects of applicants' reactions to cognitive ability tests and an assessment center. *Personnel Psychology, 47*, 715-738.

Maertz, C. P., Jr., Bauer, T. N., Mosley, D. C., Jr., Posthuma, R. A., & Campion, M. A. (2004). Do Procedural Justice Perceptions in a Selection Testing Context Predict Applicant Attraction and Intention Toward the Organization? *Journal of Applied Social Psychology, 34*, 125-145.

McCarthy, J. M., Bauer, T. N., Truxillo, D. M., Anderson, N. R., Costa, A. C., & Ahmed, S. M. (2017). Applicant Perspectives During Selection: A review Addressing "So What?", "What's New?" and "Where to Next?" *Journal of Management.* Advance online publication.

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. L. (2007). Situational judgment tests, response instructions and validity: A meta-analysis. *Personnel Psychology, 60*, 63-85.

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: a clarification of the literature. *Journal of Applied Psychology, 80,* 730-740.

McDaniel, M. A., Psotka, J., Legree, P. J., Yost, A. P., & Weekley, J. A. (2011). Toward an understanding of situational judgment item validity and group differences. *Journal of Applied Psychology, 96,* 327-336.

McFarland, L. A. (2003). Warning Against Faking on a Personality Test: Effects on Applicant Reactions and Personality Test Scores. *International Journal of Selection and Assessment, 11,* 265-276.

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17,* 437-455.

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75,* 640-647.

Muthén, L. K., & Muthén, B. O. (1998-2015). *Mplus User's Guide*. Seventh Edition. Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling, 4*, 599−620.

Olson-Buchanan, J. B., Drasgow, F., Moberg, P. J., Mead, A. D., Keenan, P. A., & Donovan, M. A. (1998). Interactive video assessment of conflict resolution skills. *Personnel Psychology, 51,* 1-24.

O'Reilly, K. A. (2010). *Service undone: A grounded theory of strategically constructed silos and their impact on customer-company interactions from the perspective of retail employees*. (Unpublished Doctoral dissertation).Utah State University, Utah.

O'Reilly, K., Paper, D., & Marx, S. (2012). Demystifying grounded theory for business research. *Organizational Research Methods, 15,* 247-262.

Ployhart, R. E., & Holtz, B. C. (2008). The diversity–validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selec*tion. Personnel Psychology, 61*, 153-172.

Ployhart, R. E., & MacKenzie, W. I., Jr. (2011). Situational judgment tests: A critical review and agenda for the future. In S. Zedeck (Ed.), *APA Handbooks in Psychology. APA handbook of industrial and organizational psychology, Vol. 2. Selecting and developing members for the organization* (pp. 237-252).

Ployhart, R. E., & Ryan, A. M. (1998). Applicants' reactions to the fairness of selection procedures: The effects of positive rule violations and time of measurement. *Journal of Applied Psychology, 83*, 3-16.

Ployhart, R. E., & Ward, A. -K. (2013). Situational judgment measures. In Geisinger, K. F., Bracken, B. A, Carlson, J. F., Hansen, J.-I. C., Kuncel, N. R., Reise, S. P., & Rodriguez, M. C. (Eds). *APA handbook of testing and assessment in psychology, Vol. 1: Test theory and testing and assessment in industrial and organizational psychology. APA handbooks in psychology*, (pp. 551-564). Washington, DC, US: American Psychological Association, xxix, 71.

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescription. *Multivariate Behavioral Research, 42,* 185–227.

Putka, D. J., Le, H., McCloy, R. A., & Diaz, T. (2008). Ill-structured measurement designs in organizational research: Implications for estimating interrater reliability. *Journal of Applied Psychology, 93*, 959-981.

Richman-Hirsch, W. L., Olson-Buchanan, J. B., & Drasgow, F. (2000). Examining the impact of administration medium on examinee perceptions and attitudes. *Journal of Applied Psychology, 85*, 880-887.

Roch, S. G., Mishra, V., & Trombini, E. 2014. Does selection measure scoring influence motivation: One size fits all? *International Journal of Selection and Assessment, 22*, 23-38.

Ryan, A. M., & Huth, M. (2008). Not much more than platitudes? A critical look at the utility of applicant reactions research. *Human Resource Management Review, 18,* 119-132.

Ryan, A. M., & Ployhart, R. E. (2000). Applicants' perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management, 26,* 565-606.

Ryan, A. M., Sacco, J. M., McFarland, L. A., & Kriska, S. D. 2000. Applicant self-selection: Correlates of withdrawal from a multiple hurdle process. *Journal of Applied Psychology, 85,* 163-179.

Sanchez, R.J., Truxillo, D.M., & Bauer, T. N. (2000). Development and examination of an expectancy-based measure of test-taking motivation. *Journal of Applied Psychology, 85,* 739-750.

Schleicher, D. J., Venkataramani, V., Morgeson, F. P., & Campion, M. A. (2006). So you didn't get the job...Now what do you think? Examining opportunity-to-perform fairness perceptions. *Personnel Psychology, 59,* 559-590.

Schmit, M. J., & Ryan, A. M. (1997). Applicant withdrawal: The role of test-taking attitudes and racial differences. *Personnel Psychology, 50,* 855-876.

Siskind, A. D. (2012). *Effect of fidelity on computerized simulation assessment outcomes* (Unpublished doctoral dissertation). Illinois Institute of Technology, IL.

Smither, J. W., Reilly, R. R., Millsap, R. E., Pearlman, K., & Stoffey, R. W. (1993). Applicant reactions to selection procedures. *Personnel Psychology, 46,* 49-76.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Thousand Oaks, CA: Sage.

Sydell, E., Ferrell, J., Carpenter, J., Frost, C., & Brodbeck, C. C. (2013). Simulation scoring. In M. Fetzer & K. Tuzinski (Eds.), *Simulations for personnel selection* (pp. 83-107).

Tonidandel, S., & Quiñones, M. (2000). Psychological reactions to adaptive testing. *International Journal of Selection and Assessment, 8,* 7-15.

Tonidandel, S., Quinones, M. A., & Adams, A. A. (2002). Computer-adaptive testing: The impact of test characteristics on perceived performance and test takers' reactions. *Journal of Applied Psychology, 87,* 320-332.

Truxillo, D. M., Bauer, T. N., Campion, M. A., & Paronto, M. E. (2006). A field study of the role of big five personality in applicant perceptions of selection fairness, self, and the hiring organization. *International Journal of Selection and Assessment, 14,* 269-277.

Truxillo, D. M., Bauer, T. N., & McCarthy, J. M. (2015). Applicant Fairness Reactions to the Selection Process in R. S. Cropanzano & M. L. Ambrose (Eds.), The Oxford Handbook of Justice in the Workplace (pp. 621 – 640).

Tuzinski, K. (2013). Simulations for personnel selection: An introduction. In M. Fetzer & K. Tuzinski (Eds.), *Simulations for personnel selection* (pp. 1-13).

Wang, C., Lo, Y.-Y., Xu, Y., Wang, Y., & Porfeli, E. (2007). Constructing the search for a job in academia from the perspectives of self-regulated learning strategies and social cognitive career theory. *Journal of Vocational Behavior, 70,* 574-589.

Weekley, J. A., Hawkes, B., Guenole, N., & Ployhart, R. E. (2015). Low-Fidelity Simulations. *Annual Review of Organizational Psychology and Organizational Behavior, 2,* 295-322.

Weekley, J. A., & Ployhart, R. E. (2005). Situational Judgment: Antecedents and Relationships with Performance. *Human Performance, 18,* 81-104.

Weekley, J. A., & Ployhart, R. E. (2006). An introduction to situational judgment testing. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment test: Theory, measurement and application* (pp. 1-10). Mahwah, NJ: Lawrence Erlbaum Associates.

Whitman, D. S., Kraus, E., & Van Rooy, D. L. (2014). Emotional intelligence among black and white job applicants: Examining differences in test performance and test reactions. *International Journal of Selection and Assessment, 22,* 199-210.

Wilhelmy, A., Kleinmann, M., Konig, C. J., Melchers, K. G., and Truxillo, D.M. (2016). How and why do interviewers try to make impressions on applicants? A qualitative study. *Journal of Applied Psychology, 101,* 313-332.

Viswesvaran, C., & Ones, D. S. (2004). Importance of Perceived Personnel Selection System Fairness Determinants: Relations with Demographic, Personality, and Job Characteristics. International *Journal of Selection and Assessment, 12,* 172-186.

APPENDIX A

PILOT STUDY INTERVIEW QUESTIONS

1. Why have an SJT branch as opposed to use a simple, linear or nonlinear SJT?
2. How does the branching work?
3. Does branching create any limitations?
4. How are BSJTs scored?
5. How are the psychometric properties of the test assessed?
6. Is adverse impact higher or lower compared to traditional SJTs?
7. Are BSJTs more or less cognitively loaded than traditional SJTs?
8. Does branching impact the underlying construct being measured?
9. What constructs have you measured using BSJTs?
10. What are the important questions that need to be asked or answered regarding BSJTS?
11. Are you aware of any literature or white papers that may help inform research on BSJTS?
12. Are there any other subject matter experts that you recommend I contact?

APPENDIX B

STUDY 1 INTERVIEW QUESTIONS

1) Can you tell me a little about your experience with BSJTs?
2) What was the motivation behind creating BSJTs for your organization/institution?
3) How does the assessment branch?
4) Does each RO branch?
5) Are the branches parallel?
6) Are the branches of equal difficulty?
7) How do you think branching impacts applicant reactions?
8) Do you create the BSJTs to have a specific impact on applicant reactions?
9) How do you measure applicant reactions to BSJTs?
10) Do you have any materials to share?

* Interview will be semi-structured and questions are likely to change as interviews are conducted.

APPENDIX C

BSJT DESCRIPTIONS

Imagine that you are applying for a job that you really want. As part of the hiring process, you take a test with the following characteristics. Please make an effort to remember these details, because you will be asked several questions about them as you complete the next several pages of survey questions. Each question is made up of a 3 or 4 sentence summary of a work-related situation. The following is an example of such a scenario:

> Recently one of your employees has had difficulty turning in assignments on time. In addition to this, the work the employee is turning in contains a large number of easily preventable errors. Previously, this individual was a model employee who always met deadlines and turned in high quality work. This employee can occasionally be sensitive to criticism. What would you do?

- Each question has four possible answers detailing different ways of responding to the scenario. You must pick one of the four answers. Your score on this test will be used to determine your suitability for the job you are applying for.
- [base branching info] The questions and answer options are not the same for all test takers. After the first item, the questions you see will depend on how you answered earlier questions.
- [narrative; contingency = 0] As you work through the test, a story develops. Each choice you make when answering determines the direction the story takes. Although there are right and wrong answers to most questions, the story will only change based upon the choices you make that don't affect how well you score. In other words, the different paths you can take are only there to tell a better story.
- [performance; contingency = 1] As you work through the test, a story develops. Each choice you make when answering determines the direction the story takes. But in addition to the story, most questions have right and wrong answers. As you answer questions, the test will change based on whether you are getting questions right or wrong. In other words, the different paths you can take are chosen based on how well you're scoring.

- [parallel; parallelism = 1] In general, the test is designed so that the differences between the questions you see and the questions other test takers see are as small as possible.
- [non-parallel; parallelism = 0] In general, the test is designed so that you could see completely different questions from other test takers.
- [looping; looping = 1] Once you've answered a question, you can return to it or any other question you previously answered and change it in order to go down a different path. However, your only your first response will count toward your score.
- [linear; looping = 0] Once you've answered a question, you will not be allowed to return to that question or any previous questions.

APPENDIX D

MEASUREMENT SCALES

Face Validity

Please rate how well each item describes you on a scale from 1 to 5.

(1 = very unlike me, 2 = somewhat unlike me, 3 = neutral, 4 = somewhat like me, 5 = very like me)

1) I would not understand what the test had to do with the job.
2) I could not see any relationship between the test and what would be required on the job.
3) It would be obvious to anyone that the test is related to the job.
4) The actual content of the test would be clearly related to the job.
5) There was no real connection between the test that I read about and the job.

Selection Procedural Justice Scale (SPJS) - Consistency of Administration

Strongly disagree = 1, Disagree = 2, Neither agree nor disagree = 3, Agree = 4, Strongly agree = 5

1) The test would be administered to all applicants in the same way.
2) There would be no differences in the way the test was administered to different applicants.
3) Test administrators would make no distinction in how they treated applicants.

Selection Procedural Justice Scale (SPJS) – Opportunity to Perform

Strongly disagree = 1, Disagree = 2, Neither agree nor disagree = 3, Agree = 4, Strongly agree = 5

1) I could really show my skills and abilities through this test.
2) This test would allow me to show what my job skills are.
3) This test would give applicants the opportunity to show what they can really do.
4) I would be able to show what I can do on this test.

Motivation

Disagree = 1, Agree = 7

Please rate how well each item describes you on a scale from 1 to 7.

1) Doing well on this test would be important to me.

2) I would want to do well on this test.
3) I would try my best on this test.
4) I would try to do the very best I could do on this test.
5) While taking this test, I would concentrate and try to do well.
6) I would want to be among the top scorers on this test.
7) I would push myself to work hard on this test.
8) I would be extremely motivated to do well on this test.
9) I wouldn't put much effort into this test.*
10) I just wouldn't care how I did on this test.*

*negatively scored

Procedural Justice

To a small extent = 1, to a large extent = 5

The following items refer to the procedures used to arrive at your (outcome). To what extent:

1) Would you be able to express your views and feelings during those procedures?
2) Would you have influence over the score arrived at by those procedures?
3) Would those procedures be applied consistently?
4) Would those procedures be free of bias?
5) Would those procedures be based on accurate information?
6) Would you be able to appeal the score arrived at by those procedures?
7) Would those procedures uphold ethical and moral standards?

Test Attitude

Strongly disagree = 1, Disagree = 2, Neither agree nor disagree = 3, Agree = 4, Strongly agree = 5

1) I would be in favor of having to take this type of test.
2) I would not like to take this type of test.

APPENDIX E

MANIPULATION CHECK

Based on the test description, choose the true statement:

Performance/Narrative Contingency

a) the test would present the same items in the same order to all candidates^
b) if you worked through the test, a story would unfold but your score wouldn't be impacted*
c) your score on this test would not impact your chances of obtaining the hypothetical job
d) the test would get harder as you answered questions correctly+

^ = correct answer for control condition

* = correct answer for narrative contingency

+ = correct answer for performance contingency

Parallel/Non-parallel

a) the test would present the same items in the same order to all candidates^
b) there would be very little difference between the items you would see and the items other candidates would see*
c) there could be large differences in the test content you see and the test content other candidates see+
d) you could pick more than one response for each question

^ = correct answer for control condition

* = correct answer for parallel

+ = correct answer for non-parallel

Looping/Linear

a) each question has six (6) possible answers
b) you could return to any item or question you had previously answered*
c) the number of answers options would vary from question to question
d) once you answered a question, you would not be able not return to it^+

^ = correct answer for control condition

* = correct answer for looping

+ = correct answer for linear

VITA

Craig Matthew Reddock
Department of Psychology
250 Mills Godwin Building
Old Dominion University
Norfolk, VA 23529-0267

Education

Old Dominion University (Expected Graduation May 2017)
Doctor of Philosophy – Industrial/Organizational Psychology
The University of Tennessee at Chattanooga
Master of Science – Industrial/Organizational Psychology
Oglethorpe University, Atlanta, GA
Bachelor of Arts – Psychology

Applied Experience

Associate Validation Consultant, CEB, April 2015 – Present
Research Scientist, CEB, January 2013 – March 2015
Science and Measurement Director, Logi-Serve, LLC, June 2011 – December 2012
Research Associate, Logi-Serve, LLC, May 2009 – June 2011
Research Assistant, Old Dominion University, August 2009 – December 2012

Publications and Presentations

Reddock, C.M., & Landers, R. N. (2012). Human resources. In *Encyclopedia of New Venture Management*. Thousand Oaks, CA: SAGE Publications.

Biderman, M. D., & Reddock, C. M. (2012). The relationship of scale reliability and validity to respondent inconsistency. *Personality and Individual Differences*.

Reddock, C. M., Biderman, M. D., & Nguyen, N. T. (2011). The relationship of reliability and validity of personality tests to frame-of-reference instructions and within-person inconsistency. *International Journal of Selection and Assessment*, 19(2), 119-131.

Landers, R. N., Reddock, C. M., Cavanaugh, K. J., & Proaps, A. B. (2014, May). Talent assessment using mobile devices. In T. Kantrowitz & C. M. Reddock (Chairs), *Shaping the future of mobile assessment: Research and practice update*. Symposium presented at the 29th Annual Conference of the Society for Industrial and Organizational Psychology, Honolulu, HI.

Reddock, C. M., & Coughlin, C.R. (2014, May). Biodata Item Generation: Using Cloning to Generate Multiple Tests. In C. M. Reddock & A. R. Schwall (Chairs), *New Developments in Biodata Research and Practice*. Symposium presented at the 29thAnnual Conference of the Society for Industrial and Organizational Psychology, Honolulu, HI.

Landers, R. N., Reddock, C. M., & Mogan, T. (2012, April). Learner control: Update and extension of Kraiger and Jerden (2007). In T. S. Behrend (Chair), R.N. Landers (Chair), & K. Kraiger (Discussant), *Building a science of learner control in training: Current perspectives.* Symposium presented at the 27th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.