

# A theory of branched situational judgment tests and their applicant reactions

Branched  
situational  
judgment tests

255

Craig M. Reddock  
*SHL, Atlanta, Georgia, USA, and*  
Elena M. Auer and Richard N. Landers  
*University of Minnesota, Minneapolis, Minnesota, USA*

Received 4 October 2018  
Revised 6 May 2019  
11 June 2019  
Accepted 18 December 2019

## Abstract

**Purpose** – Branched situational judgment tests (BSJTs) are an increasingly common employee selection method, yet there is no theory and very little empirical work explaining the designs and impacts of branching. To encourage additional research on BSJTs, and to provide practitioners with a common language to describe their current and future practices, we sought to develop a theory of BSJTs.

**Design/methodology/approach** – Given the absence of theory on branching, we utilized a ground theory qualitative research design, conducting interviews with 25 BSJT practitioner subject matter experts.

**Findings** – Our final theory consists of three components: (1) a taxonomy of BSJT branching features (contingency, parallelism, convergence, and looping) and options within those features (which vary), (2) a causal theoretical model describing impacts of branching in general on applicant reactions via proximal effects on face validity, and (3) a causal theoretical model describing impacts on applicant reactions among branching designs via proximal effects on consistency of administration and opportunity to perform.

**Originality/value** – Our work provides the first theoretical foundation on which future confirmatory research in the BSJT domain can be built. It also gives both researchers and practitioners a common language for describing branching features and their options. Finally, it reveals BSJTs as the results of a complex set of interrelated design features, discouraging the oversimplified contrasting of “branching” vs “not branching.”

**Keywords** Qualitative research, Selection, Grounded theory

**Paper type** Research paper

## Introduction

Branched situational judgment tests (BSJTs) are a relatively recent innovation in personnel selection. Large consulting companies now offer BSJTs as commercial products (see [AON, n.d.](#); [HumRRO, n.d.](#)), delivered with questions in static text, interactive video, and animated formats (e.g., [Lievens and Sackett, 2006](#)). The reasons reported for this increase in demand vary among practitioners, but generally include market demand for gamified assessments ([Armstrong et al., 2015](#)) and specific benefits commonly associated with gamification, such as improved applicant engagement, face validity, and test security. This assessment type first appeared in the scholarly literature when [Olson-Buchanan et al. \(1998\)](#) described the creation and validation of a conflict resolution BSJT in which participant responses to questions on a video SJT led different participants to different videos. Since then, however, there has been little research on this type of SJT. The only research we could identify comparing the impact of branching on outcomes of interest versus non-branching SJTs was one study that found that video-based BSJTs were viewed more favorably than text-based BSJTs and video-based



© Craig M. Reddock, Elena M. Auer and Richard N. Landers. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at: <http://creativecommons.org/licences/by/4.0/legalcode>

Journal of Managerial Psychology  
Vol. 35 No. 4, 2020  
pp. 255-270  
Emerald Publishing Limited  
0268-3946  
DOI 10.1108/JMP-10-2018-0434

SJTs (Kanning *et al.*, 2006). We identified only three published articles focusing on branching across the entire SJT literature. Thus, branching emerged for us as a key research concern; specifically, we sought to develop a theory to address a practical question: how should SJT branches be designed?

To do so, the present study adopted a qualitative, grounded theory approach to generate a theory of branching in SJTs from the judgments of BSJT subject matter experts (SMEs). Given the lack of extant theory about BSJTs, grounded theory provides a useful framework from which to develop an initial theory of branching driven by the experts implementing them, a vital first step toward developing a comprehensive and empirically supported model in a triangulation mixed-methods approach (Jick, 1979). This also supports more a proactive and timely approach to a research area with rapidly evolving technology, a common problem in this research literature (Landers, 2019). Specifically, we develop this theory with two long-term goals: (1) to provide a foundation for researchers to conduct confirmatory studies of BSJTs moving forward and (2) to provide expert guidance to practitioners based upon the latest SME thinking.

### *SJTs and BSJTs*

SJTs are a measurement method that presents applicants with a scenario (i.e., a situation) and asks them how they would respond to that scenario (i.e., a judgment; Campion *et al.*, 2014). SJTs typically involve presentation of a variety of situations that applicants might encounter on the job, with each situation followed by several choices. The test-taker might then indicate which of these options they would or should do; alternatively, they might rate or rank the effectiveness of each option presented (Weekley and Ployhart, 2006). Originally intended to measure judgment in work situations (McDaniel *et al.*, 2001), SJTs of this traditional design measure a combination of work-related knowledge, skills, and abilities (Weekley *et al.*, 2015; also see Lievens and Motowidlo, 2016 and associated commentaries for an extended discussion on SJT construct measurement).

BSJTs represent a family of variations on traditional SJTs defined by differing sequences of scenarios between test-takers, which depend, at least in part, on the test-takers' prior responses. With a BSJT, one or more of the responses could lead to a completely different scenario than the others. That is, if someone chooses response *b*, the next scenario they see might involve a conversation with Human Resources; if a test-taker chooses response *d*, the next scenario might involve a conversation with an employee. Because BSJTs are interactive and respond in real-time to test-taker actions, practically speaking, they must be technologically mediated. Typically, they are delivered via a web application that tracks participant responses and displays the next question depending upon some algorithm.

We identified a total of three empirical, peer-reviewed journal articles on BSJTs (Olson-Buchanan *et al.*, 1998; Kanning *et al.*, 2006; Richman-Hirsch *et al.*, 2000). In the earliest, Olson-Buchanan *et al.* (1998) developed and validated a BSJT measuring conflict resolution skills. Scores were found to converge with manager ratings of conflict resolution as well as to predict ratings of overall performance. In the second, researchers used that same BSJT assessment to examine differences in applicant reactions to different assessment media (paper and pencil, text-based, and multi-media; Richman-Hirsch *et al.*, 2000). The multimedia version was rated as more content valid, predictively valid, job-relevant, enjoyable, and shorter, and test-takers were more satisfied. In the third and most recent, Kanning *et al.* (2006) compared candidate reactions between traditional and BSJTs in two samples. In the first sample, a concurrent validation study randomly assigned 284 police officers to one of four cells, crossing branching (labeled *interactivity*) and media (text and video). Only one effect was observed: the officers viewed the branching versions to be more useful than the non-branching versions, with neither significant interactions nor main effects of branching on emotional reaction, transparency, or job

relatedness. In the second sample, a concurrent validation study asked 82 police officers to complete six versions of the same test which varied in terms of branching, scenario medium, response medium, and question sequence, finding no consistent main effects but noting that branching was associated with improved reactions for some designs. Thus, this study provided preliminary evidence that BSJTs can have a positive impact on applicant reactions in comparison to SJTs with certain design features. Unfortunately, the precise nature of these designs and associated trade-offs was left unclear.

In reviewing these three studies, it also became evident that they all characterized branching as a single characteristic of an SJT: branched or not branched. Additionally, all three described branching with single branch points, but such simple branching is not necessarily representative of all current or possible approaches to branching. Thus, the first goal of the present study was to determine from our SMEs what approaches to branch design were actually being used and if those varying approaches could be meaningfully categorized.

*RQ1.* What approaches are taken to branch SJTs? Can these approaches be sorted into meaningful categories?

#### *Applicant reactions theory and BSJTs*

Although BSJTs are only lightly explored in the literature, the literature on the types of outcomes they might affect is better developed and serves as a useful starting point from which to build a theory of BSJT effects. Specifically, working from earlier applicant reaction models, [Hausknecht et al. \(2004\)](#) developed a more comprehensive one detailing how applicant perceptions of the selection process can affect organizational outcomes. Within this theory, effects occur primarily through two sets of causal relationships: perceived procedure characteristics affect applicant perceptions which in turn impact various practical outcomes. Empirical research has generally supported this mediation path (e.g., [Carless and Hetherington, 2011](#)).

We interpreted [Hausknecht et al.'s \(2004\)](#) work to suggest that specific selection design characteristics can only directly affect perceived procedure characteristics (e.g., consistency of administration, face validity) since no other psychological antecedents are changeable via factors under the control of the selection system designer. Thus, the present study focused on building a theory of BSJTs based upon this specific causal path within [Hausknecht et al.'s \(2004\)](#) model. In short, if branching is to successfully affect applicant reactions (or eventually, organizational outcomes), it must do so by first affecting perceived procedure characteristics. Thus, a focal concern for this part of the project was understanding how SME judgments mapped onto known and emergent perceived procedure characteristics. Extant literature on the specific branching designs that improve applicant reactions was even more limited. Thus, we considered the effects of branching design to be essentially unexplored theoretical territory, making the question a prime candidate for a study utilizing grounded theory.

*RQ2.* How does the design of a BSJT impact applicant reaction, particularly through perceived procedure characteristics?

#### *Grounded theory*

The present study used [Schonfeld and Mazzola's \(2013\)](#) approach to grounded theory, a qualitative data collection method developed by [Glaser and Strauss \(1967\)](#) that allows researchers to generate theory through the systematic collection and iterative interpretation of qualitative data ([Glaser, 2010](#)). This approach is commonly used to explore new organizational phenomena (e.g., [Wilhelmy et al., 2016](#)). There are five main tenants that guide grounded theory research: the *constant comparative method*, *theoretical coding*, *theoretical*

*sampling, theoretical saturation, and theoretical sensitivity* (O'Reilly *et al.*, 2012). The first tenant is *constant comparison*, in which data collection and data analysis occur iteratively, with each one informing the other. The second tenant is *theoretical coding*, in which data are coded into groups and categories to discern the theoretical underpinnings of phenomena under investigation. The third tenant is *theoretical sampling*, which refers to sampling guided by the data and further refinement of the concepts uncovered by those data that continues until it ceases to generate novel information. It is at this point that *theoretical saturation*, the fourth tenant, occurs, which signals to the researcher that sufficient data have been collected. Finally, *theoretical sensitivity*, the fifth tenant, refers to a researcher's ability to give meaning to data, separating relevant data from irrelevant and properly distinguishing between the two.

Procedurally speaking, studies utilizing grounded theory require multiple distinct, iterative rounds of data collection in which earlier research findings inform later methods. For example, initial interview questions are developed using content expertise and a literature review, and as data are collected, questions might be modified or expanded in order to create a richer understanding of the focal phenomenon. Interview rounds are sometimes formally distinguished, as in the present study, so that there is time to consider how tenets are being heeded and what if anything must change to ensure the approach is continued correctly. Explicitly recognizing the human biases brought to each step of this process is central and steps are taken to minimize their effects. This requires a significant investment of time and continued reflection throughout data collection. Although we have omitted numerous details on these aspects of the process in the interest of length, please note that this reflective process was a constant effort throughout the project and refined our trajectory repeatedly.

## Method

### *First sampling*

The goal of the initial data collection effort was to learn about specific BSJT beliefs and activities from an initial group of SMEs to provide direction for a second, more extensive round of interviews and other data collection. Interview questions were developed based upon the applicant reactions literature summarized above and were targeted at the purpose of BSJTs, mechanics, and psychometric properties. The specific questions, which can be seen in [Table I](#) along with response summaries, covered a wide range of topics to maximize the breadth of relevant responses.

*Participants and procedure.* Initially, four SMEs were contacted because of their previous work on BSJTs known to the present first author. Of the four, three agreed to be interviewed. A snowball method was utilized ultimately identifying nine additional SMEs, of which four responded and were interviewed. Thus, a total of seven SMEs were interviewed, all PhDs in the field of industrial-organizational psychology. Each was sent interview questions ahead of time and asked if they had any documents they could share, which produced beta test content, slide decks, and notes. All interviews were conducted by the first author, who after each interview reviewed and compiled interview notes, documented commonalities with past interviews, and reviewed additional materials provided. Additionally, interviews were recorded and transcribed.

*Analysis.* The impetus for the use of BSJTs generally involved enhancing candidate experience, with several SMEs mentioning increases in engagement via more coherent narrative and enhanced realism. Although this finding mirrored [Kanning \*et al.\*'s \(2006\)](#) study, we noted significantly greater complexity in BSJT design, as expected. SMEs varied in opinion regarding which specific types of branching would have more positive (or a more negative) impacts on applicant reactions, making this a focal point for the next round of data collection.

Interview question	Summary of SME responses
Why have an SJT branch as opposed to use a simple, linear or nonlinear SJT?	<ol style="list-style-type: none"> <li>(1) Improved candidate engagement and reactions</li> <li>(2) Better representation of real-life non-linearity (content validity)</li> <li>(3) Better representation of how decisions in real life impact subsequent events</li> <li>(4) Improved story-telling</li> <li>(5) Enhanced test security</li> <li>(6) Better construct measurement compared to traditional SJTs</li> </ol>
How does the branching work?	<ol style="list-style-type: none"> <li>(1) Radical (construct-related) vs Incidental Branching (non-construct-related)</li> <li>(2) Fact-finding sequences, where the respondent is given the opportunity to talk to certain people or look at certain documents</li> <li>(3) Branching unrelated to scoring</li> <li>(4) Branching and converging</li> </ol>
Does branching create any limitations?	<ol style="list-style-type: none"> <li>(1) Test equivalence (everyone gets a different assessment)</li> <li>(2) Technology and resources</li> <li>(3) Test security when branches are parallel (i.e., the incidental content changes but not the radical content)</li> <li>(4) Retesting candidates</li> </ol>
How is the BSJT scored?	<ol style="list-style-type: none"> <li>(1) Estimate the trait level of each response option along a certain competency, as well as the effectiveness, to form a composite</li> <li>(2) Create composites with mean centering based on a stretching algorithm that adjusts for elevation and scatter</li> <li>(3) Simple effectiveness scoring on a scale</li> <li>(4) Bayesian methods</li> <li>(5) Create distance scores, taking into account mean centering and stretching</li> <li>(6) Item response theory (IRT) or deviation scores</li> </ol>
How are the psychometric properties of the test assessed?	<ol style="list-style-type: none"> <li>(1) It's complicated</li> <li>(2) Reliability cannot be assessed via test-retest reliability, because the BSJT may not be construct-homogeneous</li> <li>(3) Putka <i>et al.'s</i> (2008) study on assessing reliability can be consulted because it provides information on assessing internal reliability for data that has a lot of empty cells</li> </ol>
Is adverse impact (AI) higher or lower compared to traditional SJTs?	<ol style="list-style-type: none"> <li>(1) There is not clear research yet</li> <li>(2) The BSJT medium (video) may reduce subgroup differences</li> <li>(3) BSJTs may have similar AI to assessment centers</li> </ol>
Are BSJTs more or less cognitively loaded than traditional SJTs?	<ol style="list-style-type: none"> <li>(1) BSJTs are not by definition more cognitively demanding but some of the constructs measured or mediums used may be</li> <li>(2) BSJTs increase cognitive demands (additional demands on working memory because test-taker must remember details of the story throughout the assessment)</li> <li>(3) BSJTs reduce cognitive demand (narrative is present throughout the test so less of the test can be devoted to the stimulus scenarios)</li> </ol>

(continued)

**Table I.**  
Interview questions  
and summarized  
responses from first  
sampling

Interview question	Summary of SME responses
Does branching impact the underlying construct being measured?	<ol style="list-style-type: none"> <li>(1) The branching/construct relationship depends on the purpose of the test (i.e., if the path is important)</li> <li>(2) Yes, branching increases realism but presents measurement issues</li> <li>(3) Yes, because different candidates may see information at different points in their respective assessments</li> </ol>
What constructs have you measured using BSJTs?	<ol style="list-style-type: none"> <li>(1) BSJTs seem to mainly be used to assess soft skills, although hard skills could be assessed</li> <li>(2) Because a BSJT involves choices, ones that involve problem solving, decision making, conflict management, judgement, relating to others, leading others, planning and organizing, knowledge tests, investigative procedures, administrative procedures, working with people, developing others, and interpersonal skills are ideal</li> <li>(3) BSJTs allow for deeper construct and more complex measurement than SJTs</li> </ol>
What are the important questions that need to be asked/answered regarding BSJTs?	<ol style="list-style-type: none"> <li>(1) Validity and reliability</li> <li>(2) The impact of the question order and timing</li> <li>(3) The impact of mean centering</li> <li>(4) The ideal balance between the number of measurement points for each competency and overall test length</li> <li>(5) The types of questions that can be appropriately asked</li> <li>(6) The ideal way of scoring</li> </ol>

Table I.

SMEs noted both positive and negative aspects of BSJTs. The most prominent positive comments regarded enhanced test-taker experience, including “people think the assessment is more engaging,” and that, like real life, “decisions have consequences.” Overall, realism was a key driver of the decision to branch. Negative comments tended to involve assessment equivalence and technological challenges. According to the SMEs, “branching means everyone gets a different assessment.” 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 would not want a scenario to end in a disaster.” Furthermore, the richer the media format, the higher the cost of creating branching scenarios.

Overall, this first phase provided rich, detailed information regarding BSJTs and guided revised questions to be asked in the next data collection effort. Specifically, it provided insight into the assumptions of BSJT SMEs in why BSJTs are being used, how they are scored, their psychometric properties, and the kinds of constructs they are used to measure. Based on the SME responses in the first phase, it became clear that BSJTs are primarily used to enhance applicant reactions, but with great variability in terms of design approaches and effects.

### *Second sampling*

Based upon the first sampling, we refined the list of questions and targeted them more tightly toward our specific research questions. Additional interviews with a wider range of SMEs were conducted focusing upon applicant reactions and what ways BSJTs were being designed to facilitate anticipated effects.

*Participants.* To more comprehensively sample BSJT subject matter experts (SMEs), first authors on every published paper since the year 2000 that contained the keyword *situational judgment test*, as well as first authors on all accepted submissions to Society for Industrial and

Organizational Psychology conferences were contacted. This approach yielded an initial list of 261 potential SMEs. Additionally, announcements were posted on two Academy of Management mailing lists, resulting in two additional SMEs. Posts were made on Twitter and two relevant LinkedIn pages but yielded no additional contacts. A snowball technique is similar to that of the initial sampling, yielding 14 additional contacts. A total of 24 interviews were ultimately conducted. Of these SJT experts, six stated they were SJT but not BSJT experts, resulting in a final sample size used to develop our BSJT theory of 18. The SMEs came from several domains, including Research and Development, Consulting, Professional Services, and academia. One SME had a master's degree, and all others had PhDs.

*Procedure.* Data were again collected from varying sources, including semi-structured interviews, archival documents, and actual BSJTs. SMEs also provided secondary resources. Ultimately, 18 interviews, 1 dissertation, and 1 book chapter were coded. Interview questions served as a starting point and additional questions were asked when appropriate. Some of these questions were also asked in later interviews, consistent with the comparative method tenant. After all interviews were completed, they were transcribed and, along with secondary materials, reviewed and coded. The first author, who conducted all interviews, believed he reached theoretical saturation around the twelfth interview, because he was no longer hearing new information in relation to the research questions. However, given the limited number of SMEs, it was a concern that certain populations of SMEs might have been missing from the first 12 interviews. Thus, all 18 BSJT experts identified were interviewed to ensure that there were no meaningful gaps.

*Analysis.* As in the first sampling, traditional grounded theory methods were adopted to analyze the data collected. In the second sampling, we utilized O'Reilly *et al.*'s (2012) approach, modified in two ways. First, we added interview recordings from which transcriptions were created. Second, we utilized a member check in which a document summarizing our final models was sent to all 18 SMEs for feedback and commentary. A multi-step process was used to code the information collected. First, the two raters independently reviewed the interview transcripts and secondary materials line-by-line and developed codes independently. Interrater agreement ranged from 76 percent to 100 percent (mean agreement = 92 percent). Second, the codes were revised in a collaborative review process (Kreiner *et al.*, 2009). Third, the coders met between interviews to discuss the codes, paying attention to evolving themes, to resolve any coding discrepancies, and to refine the data being collected and reviewed. Finally, concepts and categories developed were reviewed to determine in what ways they were related and distinct. Once the categories and concepts were finalized, the member check was conducted. Nine SMEs provided feedback at this stage, including suggested name changes to the various branching features, comments on the specific paths within the models, and suggestions for ways to clarify the model and branches. No SME suggested major changes in the descriptions or models, supporting the validity of the developed theory in relation to overall SME judgments.

## Results

### *Research question 1*

Four key branching features emerged during coding: *contingency*, *parallelism*, *convergence*, and *looping*. Table II provides a summary and prototype item designs.

Within the first feature, *contingency*, BSJTs progress between questions based upon one of two design choices: narrative progression alone or narrative progression plus test performance. A finding consistent across all SMEs was that the basic purpose of branching in a BSJT is to tell a story within the assessment. Thus, narrative flow is a fundamental characteristic of all BSJTs. However, some BSJTs also branch based on test-taker performance (i.e. test-takers choosing more effective responses see different items than

**Table II.**  
Key BSJT branching  
features and  
prototypical items

Feature	Feature description	Design options	Design option definition	Item prototypical of design option
Contingency	Relatedness of the branching (progression through the SJT) to test-taker performance	(1) Narrative Alone  (1) Narrative + Performance	(1) Narrative unfolds from desired storytelling  (1) Narrative unfolds from storytelling and performance on prior questions	(1) Question 1: Which subordinate will you speak with next? a) Jen, b) Emily, c) Tiara, d) John (2) Question 2s are identical except the name: Jen/Emily/Tiara/John wants to talk... (1) Question 1: Interpersonal issues have been plaguing your team's performance. Who will you speak with about this next? a) HR representative, b) your team members individually, c) your team in a group meeting, d) your supervisor (2) Question 1 is scored by an SME-judgment derived key, and Question 2 begins a 4-way branched storyline depending upon that answer (1) Branch 1: Jen tells you about an unpleasant interaction with a coworker this morning. How will you respond to Jen's concerns? (2) Branch 2: Tiara tells you about an unpleasant interaction with a coworker this morning. How will you respond to Tiara's concerns?
Parallelism	The degree to which different branch content measures the same construct with highly similar questions	(1) High  (1) Low	(1) Content across branches designed to be similar  (1) Content across branches varies freely	(1) Branch 1: Jen tells you about an unpleasant interaction with a coworker this morning. How will you respond to Jen's concerns? (2) Branch 2: John wants help planning out the next week's work on a specific project. How will you help him? (1) (After speaking with Jen) Which subordinate will you speak with next? a) Emily, b) Tiara, c) John (1) (After speaking with Jen) After speaking with Jen, the project's team lead calls an all-hands meeting and asks you to attend...
Convergence	The way in which branches converge to limit the otherwise infinite number of branches	(1) Parent Items  (1) Narrative Convergence	(1) Test contains multiple self-contained scenarios  (1) Test-taker is returned to shared questions through narrative design	(1) (After speaking with Jen) Which subordinate will you speak with next? a) Emily, b) Tiara, c) John (1) (After speaking with Jen) After speaking with Jen, the project's team lead calls an all-hands meeting and asks you to attend...

(continued)



Feature	Feature description	Design options	Design option definition	Item prototypical of design option
Looping	The return of the test-taker to a previous portion of the test.	(1) None (1) By Choice  (1) By Requirement (1) (Optional) Changeability	(1) Does not loop (1) Candidate can choose to return to content  (1) Candidate is forced to return to content (1) Candidate can change answers that will be scored	(1) (No prompt) (1) Is that what you expected to happen? You may now go back and give a different answer. Do you want to see what happens if you change your response? (1) You only heard from one of your subordinates. Let's go back so that you can hear from everyone (1) Please note that trying again and changing your answer will (not) affect your score on this assessment

**Note(s):** BSJT design features generally vary along continua; labels represent extremes within that feature

Table II.

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 sometimes based upon their score on previous items. If they chose more or less effective responses, then the subsequent response options reflect that. Contingency varies along a continuum from narrative alone (i.e., all branching is to drive the narrative forward) to narrative + performance (i.e., after the first question, performance always affects question order), but there are many designs in between (e.g. narrative drives the first few questions, performance is used as part of a few more, and then pure narrative is used again at the end).

The second feature, *parallelism*, refers to the degree to which the content of the different branches measures the same construct with the same general type of questions. A BSJT in which 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 vary greatly in parallelism. In a parallel BSJT design, a question might ask if the test-taker wished to follow up with Linda, Sue, George, or Bob. All test-takers might then see the same subsequent question but with the name of the person in the prompt changed to match that previous response. In a non-parallel design, leading information might be provided about each of these people such that responding to Linda is seen as consistent with high agreeableness but to Bob as low agreeableness. Subsequent questions might focus then on narrowing in upon the specific score level for that trait, similar to computer adaptive testing. Most SMEs described BSJTs between these two extremes, allowing the branches to differ non-trivially to build more engaging narratives while attempting to maintain measurement equivalence of the overall test. Like contingency, parallelism varies along a continuum, from highly parallel (i.e., test varies very little between branches) to not-at-all parallel (i.e. test branches diverge in content and measurement completely) with many middle grounds.

The third feature, *convergence*, allows test developers a way to limit the scope and complexity of branching by bringing all test takers back to shared questions at pre-determined points in the narrative. Specifically, convergence gives greater flexibility in balancing assessment length with narrative complexity. There are two primary ways BSJTs converge: parent items and narrative 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. This branching can continue, or it can end, leading to another parent item. Thus, BSJTs containing parent items can be conceptualized as a tree containing a collection of items. After the test-taker proceeds to the end of a branching path within that tree, a new scenario is presented, restarting all test-takers at the beginning of a new tree. Narrative convergence occurs when a test-taker is brought through certain critical narrative story beats via a planned progression, typically to ensure that they encounter vital information or a vital measurement point before proceeding further. 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 does not get the relevant information. To ensure that the second test-taker is not disadvantaged by their choices, the test-taker is directed into a subsequent narrative path where they are required to interact with Joe to ensure that they see that information. All BSJTs described by SMEs utilized some convergence, so these design options should be considered distinct types. Any particular BSJT might utilize parent items, narrative convergence, or both.

The fourth and final feature, *looping*, occurs when a test-taker can revisit a portion of the test. Looping is the most complex feature, as it can be operationalized in several different ways. Generally speaking, looping design options take one of three forms: none, by choice, or by requirement. Looping by choice 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 given the opportunity to explore alternate paths earlier in the assessment or as part of the narrative. Looping by requirement refers to designs in which the candidate is pushed back to a certain point in the test. This feature is most typically used to ensure that test-takers encounter certain information or scenarios that they may have missed because of the initial branch they experienced. Both looping by choice and looping by requirement enable an optional design characteristic, the ability to change the answers that will be scored, sometimes called “saving” by SMEs. Looping by itself only enables test-takers to experience alternative branches; changeability allows test-takers to alter their responses in the eventual scoring of the test.

### *Research question 2*

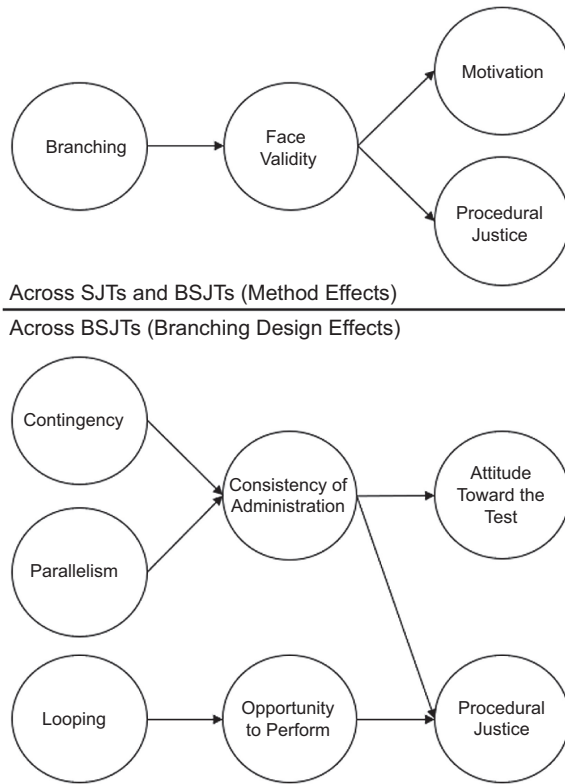
In identifying the diverse ways that branching impacts applicant reactions, several themes emerged reinforcing that BSJTs impact applicant reactions via perceived procedure characteristics and applicant perceptions. The three most commonly described perceived procedure characteristics were *face validity*, *consistency of administration*, and *opportunity to perform*. SMEs universally viewed BSJTs as more face valid because of the way that they more realistically capture the consequences of decisions and demands of the actual job. BSJTs were also viewed to impact a test-taker’s perceived opportunity to perform because they adapt to test-taker choices, giving the test-taker the sense that they can better demonstrate their true skills. The impact of consistency of administration on applicant reactions was described as both positive and 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. However, test-takers might not like the fact that the test taking experience is not standardized.

According to [Hausknecht et al. \(2004\)](#), these perceived procedure characteristics should impact applicant perceptions, and three such perceptions emerged from SMEs as focal for BSJTs: *test motivation*, *attitude toward the test*, and *perceptions of procedural justice*. BSJTs were viewed to increase test motivation because the story-like nature of the test makes candidates feel more immersed in the test. This feeling of immersion was believed to lead to increased test-taking motivation. BSJTs were also suspected to be viewed more positively as a selection measure because of the interactive feel. Finally, BSJTs were presented as 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 is driven by their own choices and not some predetermined path. Fundamentally, it was believed that because the test more closely resembled real life judgments, these outcomes would be improved.

### *Development of theoretical causal models*

As we considered the results to this point, we realized from our coding process that there were two very different sets of concerns; the general idea of branching was believed by SMEs to lead to specific changes in perceptions, mostly related to face validity. However, within branching, there were strong beliefs as to the specific values and risks associated with particular options within an overall branching design. This led us to develop two theoretical models, one comparing SJTs with BSJTs, and the other exploring design options among BSJTs only.

*Model comparing SJTs and BSJTs.* As shown in the top of [Figure 1](#), all perceived procedure characteristics described by SMEs as common to all BSJTs were categorized as face validity. In short, BSJTs deliver a narrative that adapts itself according to test-taker decisions, and this interactivity was believed by SMEs to be perceived as more reflective of actual work, where decisions have consequences that lead to more decisions. The enhanced face validity of



**Figure 1.**  
Theoretical models of  
the impact of  
branching on applicant  
reactions among SJTs

narrative branching is theorized to lead to increased test motivation and procedural justice perceptions, the two applicant perceptions most centrally described by SMEs in relation to branching in general.

*Model comparing BSJT designs.* As shown in the bottom of Figure 1, three BSJT features were modeled with distinct theoretical effects related to BSJT design. Convergence was not included in this model because SMEs described its value as saving test development time or to improving measurement characteristics. Because convergence is mostly invisible to actual test-takers, it was not believed to impact applicant reactions. Additionally, the optional feature of looping, changeability, was not included because it can only be used when looping is present and would have necessitated a third, more fine-grained model.

In developing the relationships with contingency depicted in Figure 1, we determined that narrative plus performance contingency is conceptually similar to how computer adaptive tests adapt to test-taker performance in that the skill level of the test-taker drives the selection of future test questions. In that literature, differences between perceived ability level and the difficulty of the specific items being delivered can create an incongruity between perceived performance and actual performance (Tonidandel *et al.*, 2002). Specifically, high ability test-takers may note that despite their ability level, the test is difficult, which occurs due to the standard mechanics of adaptive tests. In such situations, 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 similarly likely to cause people to attend to the fact that the test is differentially difficult 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 attitudes toward the test and on perceptions of procedural justice (Hausknecht *et al.*, 2004) depending on their subjective evaluation of the lack of consistency of administration.

In developing relationships with parallelism, we similarly reasoned that branches can vary greatly in the content they present to test-takers, thus making consistency of administration the most salient procedure characteristic related to this feature as well. In some instances, two branches may be virtually identical with only a few minor changes 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. 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).

Finally, in developing theory related to looping, we reasoned that providing the ability for test-takers to explore additional choices or even undo previous choices allows them to 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. Even in BSJTs without changeability, review of earlier scenarios could be used as a learning opportunity for test-takers to perform better in later scenarios. This enhanced perception of opportunity to perform is in turn most likely to lead to more positive perceptions of procedural justice given prior links between these constructs (Hausknecht *et al.*, 2004).

## Discussion

This study utilized grounded theory to develop an initial theory of BSJTs. This theory includes three major components: (1) a taxonomy of BSJT branching features and options within those features, (2) a causal theoretical model describing impacts of branching in general on applicant reactions, and (3) a causal theoretical model describing impacts on applicant reactions among branching designs. Our branching taxonomy contains four features of branching: contingency, parallelism, convergence, and looping. Together, these features paint a rich and complex landscape of BSJT design currently undescribed in the research literature, revealing many potential avenues for future research of significant theoretical and practical value. “Branching or not” is an oversimplification of this landscape and should be avoided as an organizational framework in BSJT research and practice.

The second major component was our causal model of branching. Although SMEs discussed a wide array of applicant reactions, common themes emerged across all the types of BSJTs described. First, the perceived procedure characteristics of face validity emerged as the most central construct targeted by all BSJT designs, as did the more distal applicant perceptions of test motivation, and perceptions of procedural justice. The model describing this is presented in Figure 1. Specifically, the model states that branching in general is likely to impact both motivation and perceived procedural justice by improving the face validity of the test. At their core, BSJTs tell a story over the course of the assessment, and this story is guided by the test-taker’s choices. SJTs are already viewed relatively favorably among selection methods, and branching has significant potential to improve them further.

Our third and final contribution was our causal model of branching design, also shown in Figure 1. Contingency branching was theorized to decrease perceived procedural justice and

test attitude by decreasing perceptions of consistency of administration with a narrative plus performance design. This suggests that contingency would only be useful to practitioners when some other gain offsets this cost, such as improved measurement characteristics. Non-parallel branching was also theorized to impact perceived procedural justice and test attitude by decreasing consistency of administration. Similarly, this suggests that non-parallelism should only be used if it creates other gains that offset this negative effect. Finally, inclusion of looping was theorized to increase perceived procedural justice by increasing opportunity to perform. This suggests that looping by choice is likely to improve applicant reactions, although the importance or impact of changeability is left to future research.

Importantly, this research primarily explored the effects of branching on applicant reactions. This was largely driven by practitioner interest; specifically, most SMEs that were consulted focused upon applicant reactions as the primary goal of investing in branching SJTs. Despite this focus, three additional research directions emerged as important next areas of investigation. First, the effects on test validity likely vary by branching strategy; the most likely to be problematic, as noted earlier, is non-parallelism, because it can create inconsistency in which constructs are assessed across persons, or inconsistency in which questions are used to assess the same constructs. There may be clever ways to design BSJTs to avoid this validity problem by using item-response theory as the basis for measurement, but this issue is wholly unexplored in the literature. Conversely, comparisons of the validity of branching vs. non-branching SJTs are generally missing from the literature, yet some existing research suggests that high-quality branching, by adding meaningful and realistic context to scenario prompts, may improve validity (Krumm *et al.*, 2015). Second, test bias and fairness may vary by BSJT design strategy, particularly if some narrative paths are differentially attractive by race, sex, or membership in any other protected class. To understand this will require a better understanding of narrative design than is commonly found in the assessment literature, potentially borrowing from the gamification literature on narrativization (Armstrong and Landers, 2017). Third, utility remains a major question for BSJTs. Different branching design strategies require different investments of time and expertise, making them differentially valuable. If the sole benefit of building a branching system is improvement of applicant reactions, then the investment of resources to achieve those increased reactions should be considered carefully. For example, the design of BSJTs incorporating low parallelism will likely require a greater investment of time to ensure their psychometric rigor remains unharmed than those with high parallelism; as such, under what conditions is low parallelism a good investment?

Although SJTs have been in use for almost a century (McDaniel *et al.*, 2001), BSJTs are a recent innovation in employee selection. The present study sought to identify the value of implementing branching in SJTs, which often carries a significant development cost. Through grounded theory methodology, we developed a taxonomy of branching features and two theoretical causal models explaining why both researchers and practitioners should care about branching primarily in terms of applicant reactions. By presenting this theory to the research community, we hope to stimulate further research and thoughtful practice about BSJTs not just in the area of applicant reactions but across all areas of concern, such as validity, fairness and bias, and utility. More broadly, we hope that this research will inspire others to explore new selection technologies in a timely fashion and using a broader array of research methods than is currently common. If we restrict ourselves to merely passively observing new selection technologies after they are fully developed with confirmatory, quantitative techniques, waiting until the technologies have stabilized instead of providing recommendations as they are created, the research literature will eventually become irrelevant to practice. It is through co-creation of knowledge between academia and industry that the highest quality technology-driven assessments will be both developed and understood.

---

## References

- AON (n.d.), "chatAssess fact sheet", available at: <https://insights.humancapital.aon.com/assessments-situational-judgment/chatassess-fact-sheet>.
- Armstrong, M.B. and Landers, R.N. (2017), "An evaluation of gamified training: using narrative to improve reactions and learning", *Simulation and Gaming*, Vol. 48, pp. 513-538.
- Armstrong, M.B., Landers, R.N. and Collmus, A.B. (2015), "Gamifying recruitment, selection, training, and performance management: game-thinking in human resource management", in Davis, D. and Gangadharbatla, H. (Eds), *Handbook of Research on Trends in Gamification*, Information Science Reference, Hershey, PA, pp. 140-165.
- Campion, M.C., Ployhart, R.E. and MacKenzie, W.I. (2014), "The state of research on situational judgment tests: a content analysis and directions for future research", *Human Performance*, Vol. 27, pp. 283-310.
- Carless, S.A. and Hetherington, K. (2011), "Understanding the applicant recruitment experience: does timeliness matter?", *International Journal of Selection and Assessment*, Vol. 19, pp. 105-108.
- Glaser, B.G. (2010), "The future of grounded theory", *The Grounded theory Review: An International Journal*, Vol. 9, pp. 1-14.
- Glaser, B.G. and Strauss, A.L. (1967), *The Discovery of Grounded Theory: Strategies for Qualitative Research*, Aldine, New York, NY.
- Hausknecht, J.P., Day, D.V. and Thomas, S.C. (2004), "Applicant reactions to selection procedures: an updated model and meta-analysis", *Personnel Psychology*, Vol. 57, pp. 639-683.
- HumRRO (n.d.), "Update on rich-media simulations", available at: <https://www.humrro.org/corpsite/article/interactive-rich-media-simulations/>.
- Jick, T.D. (1979), "Mixing qualitative and quantitative methods: triangulation in action", *Administrative Science Quarterly*, Vol. 24, pp. 602-611.
- Kanning, U.P., Grewe, K., Hollenberg, S. and Hadouch, M. (2006), "From the subjects' point of view: reactions to different types of situational judgment items", *European Journal of Psychological Assessment*, Vol. 22, pp. 168-176.
- Kreiner, G.E., Hollensbe, E.C. and Sheep, M.L. (2009), "Balancing borders and bridges: negotiating the work-home interface via boundary work tactics", *Academy of Management Journal*, Vol. 52, pp. 704-730.
- Krumm, S., Lievens, F., Huffmeier, J., Lipnevich, A.A., Bendels, H. and Hertel, G. (2015), "How "situational" is judgment in situational judgment tests?", *Journal of Applied Psychology*, Vol. 100, pp. 464-480.
- Landers, R.N. (2019), "The existential threats to I-O psychology highlighted by rapid technological change", in Landers, R.N. (Ed.), *Cambridge Handbook of Technology and Employee Behavior*, Cambridge University Press, New York, NY, pp. 3-21.
- Lievens, F. and Motowidlo, S. (2016), "Situational judgment tests: from measures of situational judgment to measures of general domain knowledge", *Industrial and Organizational Psychology*, Vol. 9, pp. 3-22.
- Lievens, F. and Sackett, P.R. (2006), "Video-based versus written situational judgment tests: a comparison in terms of predictive validity", *Journal of Applied Psychology*, Vol. 91, pp. 1181-1188.
- McDaniel, M.A., Morgeson, F.P., Finnegan, E.B., Campion, M.A. and Braverman, E.P. (2001), "Use of situational judgment tests to predict job performance: a clarification of the literature", *Journal of Applied Psychology*, Vol. 80, pp. 730-740.
- Olson-Buchanan, J.B., Drasgow, F., Moberg, P.J., Mead, A.D., Keenan, P.A. and Donovan, M.A. (1998), "Interactive video assessment of conflict resolution skills", *Personnel Psychology*, Vol. 51, pp. 1-24.

- O'Reilly, K., Paper, D. and Marx, S. (2012), "Demystifying grounded theory for business research", *Organizational Research Methods*, Vol. 15, pp. 247-262.
- Putka, D.J., Le, H., McCloy, R.A. and Diaz, T. (2008), "Ill-structured measurement designs in organizational research: implications for estimating interrater reliability", *Journal of Applied Psychology*, Vol. 93, pp. 959-981.
- Richman-Hirsch, W.L., Olson-Buchanan, J.B. and Drasgow, F. (2000), "Examining the impact of administration medium on examinee perceptions and attitudes", *Journal of Applied Psychology*, Vol. 85, pp. 880-887.
- Schonfeld, I.S. and Mazzola, J.J. (2013), "Strengths and limitations of qualitative approaches to research in occupational health psychology", in Sinclair, R.R., Wang, M. and Tetrick, L.E. (Eds), *Research Methods in Occupational Health Psychology: Measurement, Design, and Data Analysis*, Routledge/Taylor and Francis Group, New York, NY, pp. 268-289.
- Tonidandel, S., Quiñones, M.A. and Adams, A.A. (2002), "Computer-adaptive testing: the impact of test characteristics on perceived performance and test takers' reactions", *Journal of Applied Psychology*, Vol. 87, pp. 320-332.
- Weekley, J.A., Hawkes, B., Guenole, N. and Ployhart, R.E. (2015), "Low-fidelity simulations", *Annual Review of Organizational Psychology and Organizational Behavior*, Vol. 2, pp. 295-322.
- Weekley, J.A. and Ployhart, R.E. (2006), "An introduction to situational judgment testing", in Weekley, J.A. and Ployhart, R.E. (Eds), *Situational Judgment Test: Theory, Measurement and Application*, Lawrence Erlbaum Associates, Mahwah, NJ, pp. 1-10.
- Wilhelmy, A., Kleinmann, M., König, 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*, Vol. 101, pp. 313-332.

**Corresponding author**

Richard N. Landers can be contacted at: [rlanders@umn.edu](mailto:rlanders@umn.edu)