

Singapore Management University

## Institutional Knowledge at Singapore Management University

---

Research Collection Lee Kong Chian School Of  
Business

Lee Kong Chian School of Business

---

1-2020

### Situational Judgment Tests: From low-fidelity simulations to alternative measures of personality and the person-situation interplay

Filip LIEVENS

Singapore Management University, [filiplievens@smu.edu.sg](mailto:filiplievens@smu.edu.sg)

Philipp SCHÄPERS

Follow this and additional works at: [https://ink.library.smu.edu.sg/lkcsb\\_research](https://ink.library.smu.edu.sg/lkcsb_research)



Part of the [Human Resources Management Commons](#)

---

#### Citation

LIEVENS, Filip and SCHÄPERS, Philipp. Situational Judgment Tests: From low-fidelity simulations to alternative measures of personality and the person-situation interplay. (2020). *Emerging approaches to measuring and modeling the person and situation*. Research Collection Lee Kong Chian School Of Business.

Available at: [https://ink.library.smu.edu.sg/lkcsb\\_research/6501](https://ink.library.smu.edu.sg/lkcsb_research/6501)

This Book Chapter is brought to you for free and open access by the Lee Kong Chian School of Business at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection Lee Kong Chian School Of Business by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email [liblR@smu.edu.sg](mailto:liblR@smu.edu.sg).

**Situational Judgment Tests:**

**From Low-fidelity Simulations to Alternative Measures of Personality and the Person-Situation Interplay**

Filip Lievens & Philipp Schäpers

Singapore Management University

Christoph N. Herde

Ghent University

*This is the final author version (before journal's typesetting and copyediting) of the following chapter: Lievens, F., Schäpers, P., & Herde, C.N. (in press). Situational Judgment Tests: From low-fidelity simulations to alternative measures of personality and the person-situation interplay. In D. Wood, P. Harms, S. Read, & A. Slaughter (Eds.) *Emerging approaches to measuring and modeling the person and situation*. Elsevier.*

This chapter is not the copy of record and may not exactly replicate the final, authoritative version of the chapter.

Address for correspondence: Prof. Dr. Filip Lievens, Singapore Management University, Lee Kong Chian School of Business, 50 Stamford Road, 178899, Singapore, [filiplievens@smu.edu.sg](mailto:filiplievens@smu.edu.sg)

### **Abstract**

In employment and education settings, Situational Judgment Tests (SJTs) have made strong inroads. So far, however, they are still underutilized in personality research. The objective of this chapter is to outline how SJTs might be adapted to measure personality traits, shed light onto the person-situation interplay, and stimulate research on it. We start by discussing the traditional simulation-based view on SJTs, including information on their development and research results. Next, we show how more recent versions have started to assess people's knowledge of relevant behavior related to personality traits. Finally, we specify various strategies as to how SJTs might be further adapted to shed light on the personality-situation interplay. Along these lines, we show how SJTs might be used to assess within-person variability across situations, situation-trait contingencies, proactive transactions, behavioral responses, narratives and goals, and personality disorders.

**Keywords:** Situational Judgment Tests, personality-situation interplay, simulation, implicit trait policies

### **Bio of the authors**

**Filip Lievens** is Lee Kong Chian Professor of Human Resources at the Lee Kong Chian School of Business of Singapore Management University. He is also an Honorary Professor at the University of Cape Town and a visiting professor at Ghent University, Belgium. In 1999, he obtained his PhD at Ghent University, Belgium.

His main interests deal with talent acquisition, talent assessment, and adaptability. As an internationally recognized scholar, he has published over 45 papers in top-tier journals, including the *Annual Review of Psychology*, *Journal of Applied Psychology*, *Personnel Psychology*, *Journal of Management*, *Organizational Behavior and Human Decision Processes*, *Intelligence*, and *Organizational Research Methods*.

Filip Lievens serves on the editorial board of *Journal of Applied Psychology* and *Personnel Psychology*. He was the first European winner of the Distinguished Early Career Award of the Society for Industrial and Organizational Psychology. He is also a winner of the Jeanneret Award and the Douglas Bray and Ann Howard Award, which recognize outstanding contributions in the assessment domain. He was a recipient of the Friedrich Wilhelm Bessel-Forschungspreis of the Alexander von Humboldt- Foundation and is laureate of the Royal Flemish Academy of Sciences and Arts, an award to value substantial contributions in any discipline in social sciences. He is a fellow of the Society for Industrial and Organizational Psychology (SIOP), the American Psychological Association (APA), the Association for Psychological Science (APS), and the International Association of Applied Psychology (IAAP).

**Philipp Schäpers** is currently working as a Senior Research Associate at the Lee Kong Chian School of Business of Singapore Management University. His research is situated in the field of personnel selection and psychological assessment. He is interested in the underlying working mechanisms of selection and testing procedures and his work seeks to

improve the theoretical and practical understanding of selection tools. For instance, his PhD thesis revolves around the question of how situations in Situational Judgment Tests affect their psychometric properties. Further interests are diversity management, digitalization and employer branding.

**Christoph N. Herde** is currently a PhD student at the Department of Personnel Management, Work, & Organizational Psychology at Ghent University, Belgium. His main interests deal with personnel selection with special attention to modular approaches of selection methods and simulation-based methods, such as Multiple Speed Assessments, Situational Judgment Tests, or Assessment Centers. Further interests are within-person variability in behavior and performance as well as how this variability can be captured and used to assess constructs like (interpersonal) adaptability, learning agility, or stress resilience. Other research interests are research methods and statistical modeling.

## Introduction

Situational Judgment Tests (SJTs) are popular instruments in personnel selection that assess an applicant's judgment regarding critical work-related situations (e.g., Weekley, Ployhart, & Holtz, 2006; Whetzel & McDaniel, 2009). SJTs confront applicants with short job-related situations and different response options to indicate how they would react to these incidents (McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001). Reasons for SJTs' popularity are manifold: for instance, SJTs show substantial validity for the prediction of job performance (e.g., Christian, Edwards, & Bradley, 2010; McDaniel, Hartman, Whetzel, & Grubb III, 2007), favorable applicant perceptions (Kanning, Grewe, Hollenberg, & Hadouch, 2006), provide smaller subgroup differences than many other psychological tests (e.g., cognitive ability tests; Lievens, Peeters, & Schollaert, 2008; Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004, but see also Herde, Lievens, Jackson, Shalfrooshan, & Roth, in press; Roth, Bobko, & Buster, 2013), are less fakeable than self-reports (Hooper, Cullen, & Sackett, 2006; Kasten, Freund, & Staufenbiel, 2018), and are less costly than high-fidelity simulations (e.g., assessment-centre exercises; Lievens & Patterson, 2011). Notably, SJTs have also emerged as a valid tool for assessment and selection in medical and educational settings (e.g., admission and credentialing purposes).

The growing interest in SJT research is also attested by an increasing number of papers and citations related to SJTs. Since Motowidlo, Dunnette, and Carter (1990) reinvigorated interest in SJTs, researchers have published 182 papers about SJTs. These papers have been cited 2,854 times in 1,702 articles as of December 2018. As shown in Figure 1, research interest in SJTs has constantly increased as indicated by researchers conducting and citing a growing number of SJT publications.

INSERT FIGURE 1 HERE

In personality research, scenario-based approaches like SJTs are less well known. Granted, S-R inventories (Endler, Hunt, & Rosenstein, 1962; Furnham & Jaspars, 1983) have existed for a long time but these are not job-related. Recently, there has also been growing interest in adopting SJTs for personality research (e.g., Lievens, 2017a; Mussel, Gatzka, & Hewig, 2018; Oostrom, de Vries, & de Wit, 2019). These researchers argued that personality research can benefit from scenario-based measures of personality, such as SJTs, in order to complement common approaches in personality assessment. In a special issue article, Lievens (2017a) outlined how SJTs might have great potential as alternative measures of personality and of the person-situation interplay. He argued that their use might extend from typical employment and educational settings to personality, clinical, and social psychology. Lievens pointed out that SJTs might improve research on within-person variability, personality disorders, or personality facets. Thirty-two researchers in *Personality, Social and Industrial/Organizational Psychology* commented on this article and generally expressed their interest in using SJTs in these domains.

The objective of this chapter is to outline how SJTs might be adapted to measure personality traits, shed light onto the person-situation interplay and stimulate research on it. The structure of this chapter is as follows. First, we start with a brief definition and depiction of the past history of research on SJTs. Second, we describe the traditional view of SJTs as low-fidelity simulations. Third, we describe a more recent perspective that considers SJTs as alternative measures of personality. Fourth, we illustrate how SJTs might shed light on various forms of the personality-situation interplay. **SJTs: Definition and Brief History**

### **Definition**

In Situational Judgment Tests (SJTs), candidates are presented with short job-related situational descriptions and various response options to deal with the situations (Motowidlo et al., 1990). So, SJTs consist of two main components: a situation component and a response

component. Regarding the *situation* component, candidates typically receive a brief description about the problem and some information about the people involved in the situation. Most SJTs still take the form of a written test because the scenarios are presented in a written format. In video-based or multimedia SJTs, a number of video scenarios describing a person handling a critical job-related situation are developed (McHenry & Schmitt, 1994). At a critical “moment of truth”, the scenario freezes and applicants are asked to choose among several courses of action. Thus, video-based and multimedia SJTs allow the item context to be richly portrayed, thereby increasing their stimulus fidelity (Funke & Schuler, 1998; Olson-Buchanan & Drasgow, 2006). Recently, organizations are also exploring 2D-animated, 3D-animated, and even avatar-based SJTs (Weekley, Hawkes, Guenole, & Ployhart, 2015).

In the traditional SJT literature, situation descriptions (i.e., the item stems) are described as the heart of any SJT (e.g. Campion & Ployhart, 2013; Weekley et al., 2006). SJT theory postulates that situations in SJTs provide relevant context to applicants so that they can imagine themselves in a particular scenario and apply their context-dependent knowledge to respond to it. However, recent studies (e.g., Jackson, LoPilato, Hughes, Guenole, & Shalfrooshan, 2017; Krumm et al., 2015; Schäpers, Lievens, et al., in press; Schäpers, Mussel, et al., in press) challenged this view and found that SJTs’ psychometric properties (e.g., construct saturation, SJT performance, or reliability) are not necessarily affected when situation descriptions were stripped off. For instance, Schäpers, Lievens et al. (2019) administered a video-based leadership SJT either with or without situation descriptions. The authors found that for between 47 and 88 % (uncorrected *p*-values vs. corrected for alpha inflation) of the video SJT items, it did not make a significant difference in SJT performance whether applicants worked on the version with or without scenarios. These findings underline that sufficient time and effort need to be invested in the development of the situation



description. Otherwise, SJT situation descriptions might exert only limited impact (for further details see section SJT development). The *response* component also exists in different formats. Upon reading the short scenarios, candidates might be asked to pick one response option from a list, rank the response options, or rate the effectiveness of these options. Contrary to self-report personality scales, candidates are not asked to rate themselves on the different response options. In SJTs, two instructions are typically distinguished: a knowledge-based instruction (“What should you do?”) and a behavioral tendency instruction (“What would you do?”). Knowledge-based instructions reflect maximal performance, whereas behavioral tendency instructions are related to typical performance (e.g., McDaniel et al., 2007). So, the instructions moderate the construct validity of SJT scores: SJT scores with knowledge-based instructions show higher correlations with general mental ability, whereas SJT scores with behavioral tendency instructions show higher correlations with Big Five personality (McDaniel et al., 2007). Notably, the moderating effect is particularly noticeable for Agreeableness, Conscientiousness, and Neuroticism. Furthermore, SJTs with behavioral tendency instructions are more prone to self-deception and impression management. The criterion-related validity of SJT scores is the same for both instruction formats (McDaniel et al., 2007).

Box 1 presents an example SJT item from the Teamwork-KSA Test (Stevens & Campion, 1999). The principle behind SJTs, similar to all simulations (e.g., assessment center exercises), is that applicants’ responses predict how they will behave in similar situations in the future. Commonly, work-related situations are developed from an analysis of critical incidents and are evaluated by subject-matter experts who determine what is an effective behavior for this situation at work (for an example see Weekley et al., 2006). For this item, answer B is the keyed answer because it reflects the most effective behavior (as determined by subject-matter experts). That is, these experts reasoned that disagreeable (but

routine) tasks should be fairly distributed within a team (regardless of task and position of the team members).

**Box 1. Example item from the Teamwork-KSA Test (see Stevens & Campion, 1999)**

Suppose that you find yourself in an argument with several co-workers about who should do a very disagreeable, but routine task. Which of the following would likely be the most effective way to resolve this situation?

- A. Have your supervisor decide, because this would avoid any personal bias.
- B. Arrange for a rotating schedule so everyone shares the chore. (*correct answer*)
- C. Let the workers who show up earliest choose on a first-come, first-served basis.
- D. Randomly assign a person to do the task and don't change it.

### **Brief History**

SJTs are not new inventions. Early SJT versions go back to before World War II (WWII). During WWII, military selection psychologists needed an instrument to select competent soldiers to join the armed forces. They developed a “job test” that consisted of detailed and realistic descriptions of challenging military situations. All descriptions were situations that armed forces were likely to encounter while on the job. After reading each situation, recruits were presented with several potential reactions to the given threat and they were asked which reaction they considered the most effective response (Northrop, 1989). The instrument turned out to be a success. On the one hand, it gave a realistic job preview of what was to come, thereby discouraging recruits with an unfavorable person-organization fit and lowering attrition rates for the army. On the other hand, the instrument enabled measuring recruits’ judgment skills in job-related settings, thereby significantly facilitating competent new soldier selection. This job test can be considered as one of the first SJTs.

After WWII, several similar tests were designed to capture supervisory potential (e.g., Bruce, 1974; Cardall, 1942; File, 1945; Greenberg, 1963). In 1990, Motowidlo et al. reinvigorated interest in SJTs in the scientific community and in practice. Since then, SJTs

have become attractive selection instruments for practitioners who are looking for cost-effective instruments. Their use extends beyond employment settings because SJTs have also become popular as part of (medical) admissions and credentialing testing (e.g., Lievens & Sackett, 2012; Lievens, Sackett, & Buyse, 2009; Oswald et al., 2004; Patterson, Zibarras, & Ashworth, 2016; Prasad, Showler, Schmitt, Ryan, & Nye, 2017).

### **The Traditional View: SJTs as Low- Fidelity Simulations**

#### **Underlying Theory and Rationale**

Traditionally, SJTs are seen as simulation-based instruments that are based on the behavioral consistency logic (Lievens & De Soete, 2012; Schmitt & Ostroff, 1986; Wernimont & Campbell, 1968). That is, the assumption that candidates' performance during the selection procedure will be consistent with their future performance on the job. One of the main differences between SJTs and other simulation-based assessments relates to their level of fidelity. Fidelity can be conceptualized as the extent to which a selection task mirrors related job situations (Callinan and Robertson, 2000; De Soete & Lievens, 2015). Following this definition, some simulation-based measures (e.g., work samples or role plays) can be defined as high-fidelity simulations because applicants face actual problems (replicas of job tasks). On the other side, simulation-based measures such as SJTs are regarded as *low-fidelity* simulations (Motowidlo et al., 1990; Weekley et al., 2015) because they do not require test-takers to respond to job situations in reality, but instead confront them with written or video-based descriptions of realistic job situations. This also means that SJTs are more cost-effective than high-fidelity simulations and can be more easily administered to large groups of applicants. Hence, SJTs are often used as a pre-selection ("select-out") tool, whereas high fidelity simulations are applied during later selection stages ("select-in"; De Soete & Lievens, 2015).

Using this logic, SJTs have been developed to capture domains as diverse as teamwork knowledge (McClough & Rogelberg, 2003; Morgeson, Reider, & Campion, 2005; Stevens & Campion, 1999), aviation pilot judgment (Hunter, 2003), team roles (Morgeson et al., 2005), emotion management (Blickle et al., 2009; MacCann & Roberts, 2008), call center performance (Konradt, Hertel, & Joder, 2003), proactivity (Chan & Schmitt, 2000), goal orientation (Westring et al., 2009), or academic performance (Oswald et al., 2004; Peeters & Lievens, 2005). Christian et al. (2010) developed a taxonomy to categorize these various domains. Their categorization showed that most SJTs capture leadership skills (37.5%), heterogeneous content (33.09%), and interpersonal skills (12.5%). In addition, SJTs were also developed to measure basic personality tendencies (9.56%), teamwork skills (4.4%), and job knowledge and skills (2.94%).

### **SJT Development**

According to typical SJT guidelines the development of SJTs includes three different stages (Corstjens, Lievens, & Krumm, 2017; Motowidlo et al., 1990; Ployhart & MacKenzie Jr., 2011). In the first step, test developers conduct a job analysis to generate different item stems (i.e., situation descriptions, see Flanagan, 1954). During this job-analysis, critical incidents of work situations are collected from subject matter experts (e.g., job incumbents or supervisors) that are representative of exceptionally good or exceptionally poor job-performance. Afterwards, the situations obtained are checked for level of specificity, redundancy, etc. The rationale behind this approach is to find and generate situations that are decisive for the success of the respective job.

In the second step, subject matter experts and/or (inexperienced) job incumbents generate different response options. An SJT item should contain both effective and ineffective behaviors for a respective situation. Subject matter experts are helpful for generating effective and ineffective reactions. Inexperienced employees might help to

generate a wider range of responses with different levels of effectiveness. Normally, an SJT-item contains four to five response options (Lievens et al., 2008). Furthermore, SJT developers have to decide on the response instructions (“would do” vs. “should do” instructions) and the response format. The latter can take a rate, rank, or most/least response format (for further information and effects on cognitive load and subgroup difference, see Arthur et al., 2014).

Finally, SJT developers decide on a scoring key. The following four categories are described in the SJT literature: (1) empirical, (2) theoretical, (3) rational/expert-based, and (4) hybrid scoring key. First, empirical scoring keys score response options according to their relationships with a criterion measure. Second, theoretical scoring keys follow a theoretical framework. This means that a reflected theory is used to determine (in-)effective behavior. Third, rational/expert-based scoring keys build upon subject matter experts’ ratings of the effectiveness of the different response options. Fourth, hybrid scoring keys combine at least two of the above-mentioned scoring keys. There is evidence that the scoring strategy might affect the psychometric properties of an SJT (e.g., magnitude of criterion-related validity; for further information see Bergman, Drasgow, Donovan, Henning, & Juraska, 2006), although different scoring keys also correlate often quite highly.

### **Construct-Related Validity Evidence**

Similar to other simulations, SJTs present complex job situations that tap into multiple constructs. That is, to map the requirements of a job as well as possible, a collection of different critical incidents is usually assessed. Although this approach may be conducive for predicting successful work performance, it challenges the internal psychometric analyses of SJT scores. Accordingly, SJTs show high item heterogeneity and low internal consistency. The evidence of internal construct-related validity is often lacking (Guenole, Chernyshenko,

& Weekly, 2017) because traditional SJT scores rarely break down in an assumed factor structure.

In the context of external construct-related validity, researchers extensively examined correlations between SJT scores and personality and cognitive ability (e.g., McDaniel et al., 2007). This is because personality and cognitive ability were often described as main antecedents of SJT performance (Lievens & Motowidlo, 2016). While controlling for SJT content, meta-analytic results revealed that SJTs (especially those with a behavioral tendency response instruction) correlate moderately with scores on self-report personality scales (Agreeableness .37, Conscientiousness .34, Emotional Stability .35) and moderately with cognitive abilities (.35 with knowledge instructions; .19 with behavioral tendency response instructions). In sum, the constructs that traditional SJTs measure are not always clearly identifiable. A possible explanation for this is that SJTs commonly include a wide range of different situations and response options. Therefore, a construct-driven SJT development has increased in popularity for some years now (see Guenole et al., 2017; Lievens, 2017b; for further details please see below).

### **Criterion-Related Validity Evidence**

McDaniel et al. (2007) conducted a meta-analysis of the criterion-related validity of SJT scores in employment settings. They reported a mean corrected validity of .26. Furthermore, McDaniel et al. found that SJTs had incremental value for predicting job performance over and above cognitive ability ( $R^2$  increments .03 and .05; depending on response instructions) as well as self-rated personality ( $R^2$  increments .06 and .07). So, although people's SJT scores share variance with their self-report personality ratings, they are also sufficiently different that they have additional power to predict content-relevant behavior. Consistent with this, the same meta-analysis revealed that SJTs explained about 2% extra variance over and above both general personality and cognitive ability measures. Thus,

SJTs increment can provide valuable information (especially for large organizations with large applicant pools). However, we would also like to stress the importance of not overinterpreting the results on incremental validity. In another meta-analysis, Christian et al. (2010) categorized the different domains that are captured in SJTs and compared their criterion-related validity regarding different job performance facets. The overall pattern of results showed relatively higher validities if the construct domain of the SJT matched the construct domain of the job performance facet. Finally, video-based SJTs appeared to show higher criterion-related validity than paper-pencil-based SJTs if interpersonal skills, leadership or heterogenous composites were captured. Due to this good criterion-related validity for predicting job performance (Christian et al., 2010; McDaniel et al., 2007) and favourable job applicant reactions (Kanning et al., 2006), SJTs have made strong inroads in selection in the last twenty years.

### **SJTs in the Training Context**

Although SJTs are predominately used as selection procedures, several researchers (e.g., Thornton III, Mueller-Hanson, & Rupp, 2017; Cox, Barron, Davis, & de la Garza, 2017) have argued that SJTs could also be useful for training purposes. As cost-effective simulations, SJTs might help trainees to learn how to handle critical work situations effectively. In contrast to other simulation-oriented approaches (e.g., unstructured critical incidents, role plays or case studies), SJTs offer various advantages due to their simple and cost-sensitive application. For instance, SJTs allow trainees to put themselves in many different critical incidents and to receive feedback on performance immediately. Finally, SJTs also provide the opportunity to apply and test acquired knowledge. That is, SJTs might be useful to assess training effectiveness, especially if the evaluation of job performance is unavailable. As SJTs can be easily administered via an internet link, they serve as cost-effective pre- and post-measures (Ostroff, 1991).

Cox et al. (2017) provided initial evidence that SJTs are not only worthwhile in personnel selection but can also improve learning outcomes in a training context. They developed an SJT-based training module for managing a disaster sheltering center and found that trainees who participated in an SJT-based training module showed a greater performance improvement than trainees who participated only in a traditional lecture-based training. However, research on this topic is still in its infancy and comparatively few attempts have been made to explore the usefulness of SJTs for training purposes (for first examples see Cox et al., 2017; Hauenstein, Findlay, & McDonald, 2010). Therefore, additional research is required to fully understand to what extent SJTs are useful in a training context.

### **SJTs as Alternative Measures of Personality**

In the last years, there has been a strong push to pay more attention to the constructs measured by SJTs and adopt a more construct-driven approach for developing SJTs (Guenole et al., 2017; Lievens, 2017a, 2017b; Lievens & Motowidlo, 2016; Motowidlo, Hooper, & Jackson, 2006). This has given rise to the construction of SJTs that target (personality) traits (e.g., Bledow & Frese, 2009; Christian et al., 2010; Motowidlo et al., 2006; Mussel et al., 2018; Oostrom et al., 2019). One of the main differences to the classic SJTs is that the item stems (situations) of construct-driven SJTs are typically developed to activate a specific personality trait (e.g., Agreeableness) and the different response options represent different levels of the trait being targeted. For instance, some options might reflect lower levels of Agreeableness, whereas other options indicate higher levels of Agreeableness. Recently, Guenole et al. (2017) outlined a stepwise procedure and an excellent set of recommendations for developing construct-driven SJTs that include the key features below. Table 1 presents the key differences between traditional SJTs and construct-driven SJTs (see also Lievens, 2017b).

INSERT TABLE 1 HERE



## Characteristics and Development

SJTs that measure personality traits have the following characteristics.<sup>1</sup> First, the content (item stems and item options) of construct-driven SJTs are not necessarily developed with the help of subject matter experts. Although subject matter experts can still be of help, psychologists (e.g., graduate students familiar with personality trait psychology) are typically in charge of developing the item stems on the basis of trait activation theory (Tett & Burnett, 2003). This means that the situation depicted in the item stem is deemed relevant for eliciting the construct (trait) of interest. It is also possible to rely on situational taxonomies (see Gerpott, Thielmann, & Balliet, 2019; Parrigon, Woo, Tay, & Wang, 2017; Rauthmann et al., 2014) for developing the situations in the respective item stems. Generally, using either one of these approaches will make the item stem situations somewhat shorter, less contextualized, and more unidimensional (see Tett & Guterman, 2000 for examples of trait activation situations for various constructs).

As a second key feature, the item responses of construct-driven SJTs are more unidimensional because they operationalize different levels of the targeted trait instead of reflecting different behavioral categories. To generate the response options, SJT developers typically rely on groups of subject-matter experts (psychologists) who propose and verify that the options reflect high or low levels of the traits, with results showing that this can be adequately done (see for example Motowidlo et al., 2006; Mussel et al., 2018).

Other differences are that construct-driven SJTs typically use a “What would you do?” response instruction and that a trait score is computed in the same way as in a

---

<sup>1</sup> There is conceptual overlap between construct-driven SJTs and the revealed trait technique (e.g., Costello, Wood, & Tov, 2018; Wood, Tov, & Costello, 2015). Similar to SJTs, the revealed trait technique confronts test-takers with scenarios and requests them to rate the likelihood of engaging in proposed actions. In addition, independent coders or subject matter experts rate the degree to that each action characterizes certain traits. Test-takers’ trait levels are then estimated by the correlation between the reported likelihood of engaging in certain actions and their representativeness for the respective trait. For more information, we refer to Costello et al. (2018) or Wood et al. (2015).

personality scale by averaging (or summing) people's endorsement of the response options (either high or low on the trait of interest). So, to score such SJTs, people receive higher scores when they endorse response options that are considered expressions of high levels of the specific trait targeted and vice versa. Apart from such trait scoring, another option is that people with considerable knowledge and experience with the particular contexts and subject matter involved score the effectiveness of the options ("effectiveness scoring") and that candidates receive higher scores when their answers are similar to those of the experienced/knowledgeable people.

Box 2 presents an example SJT item from Mussel et al. (2018) that aims to measure self-discipline. For this SJTs, Mussel et al focused on a single facet within each Big Five dimension (instead of broad dimension such as leadership skills). As already noted above, the different item responses reflect different levels of the construct of interest instead of being indicative of a variety of traits (as is the case in traditional SJTs).

**Box 2. Sample item from the facet self-discipline (see Mussel et al., 2018)**

You need to prepare a presentation at short-notice for an important meeting the next day. It's early evening and you're thinking about how to proceed. What would you do?

- A. I stay longer in the office and finish the presentation. (*reflects higher levels of self-discipline*)
- B. After a short break I start working on the presentation so that I can finish it quickly. Afterwards, I enjoy the rest of the evening. (*reflects higher levels of self-discipline*)
- C. I relax in front of the TV after the stressful working day. Finally, I have enough time to finish the presentation later. (*reflects lower levels of self-discipline*)
- D. I am going to enjoy the evening and improvise in the meeting. (*reflects lower levels of self-discipline*)

## **Underlying Theory**

Conceptually, construct-driven SJTs are posited to assess implicit trait policies (ITPs). ITPs are defined as inherent beliefs about the general effectiveness of actions that express traits to varying degrees (Lievens, 2017b, 2017a; Lievens & Motowidlo, 2016; Motowidlo &

Beier, 2010; Motowidlo et al., 2006). So, ITPs reflect people's procedural knowledge about relations between expressions of traits and their effectiveness in situations and people acquire them over time through general life experiences. Therefore, ITPs are viewed as characteristic adaptations instead of basic personality tendencies (see McCrae & Costa, 1996).

Implicit trait policies are not equivalent to personality traits because a person's implicit trait policy represents his or her procedural knowledge about relations between expressions of traits and their effectiveness in situations to achieve some goal. The relationship between personality traits and implicit trait policies is grounded by the mechanism of dispositional fit (Motowidlo et al., 2006). That is, people's personality traits interact with the traits expressed by the different response actions in the SJT in such a way that people who possess high levels of the trait expressed by the action believe that this action is more effective than people who have a lower standing on the trait. In addition, Motowidlo et al. (2006) argued that people who endorse SJT response options representing actions expressing specific traits will show corresponding trait-related behavior more often in real life. In other words, the notion of dispositional fit posits that by endorsing response options in SJTs people reveal something (1) about their standings on the personality traits represented by these response options, as well as (2) their likelihoods to show behavior expressing these associated traits in real life situations (Weekley et al., 2015).

### **Construct-Related and Criterion-Related Validity Evidence**

Recent research with construct-driven SJTs provides evidence that it is possible to assess specific well-defined traits (i.e., people's procedural knowledge about the effectiveness of these traits; e.g., Bledow & Frese, 2009; Motowidlo et al., 2006; Mussel et al., 2018). As one of the best examples, Mussel et al. (2018) developed an SJT for assessing five narrow traits (self-consciousness, openness to ideas, compliance, gregariousness, self-discipline). Their results showed a lot of promise for construct-driven SJTs. On average,

facets assessed by the SJT obtained a convergent validity of .59 (from .41 to .70) with corresponding self-report ratings of these facets, while discriminant validity was -.01 (from -.31 to .19). The SJT facet ratings also predicted theoretically relevant outcomes over and above the self-report ratings of the same facets. Key “ingredients of their recipe” to successfully design such a construct-driven SJT included the reliance on clearly defined personality traits instead of vague dimensions and the development of response options to reflect low versus high trait levels. In addition, Mussel et al. developed a large set of items and measured each of the narrow traits with 22 items.

In several studies, Motowidlo et al. (2006) found support for their ideas regarding the role of ITPs. For example, they examined whether ITP scores for Agreeableness correlated with (1) self-reported Agreeableness ratings on the NEO-FFI and (2) behavioral expressions of Agreeableness as observed in actual situations. This was done via mini role-plays (e.g., interactions with coworkers, subordinates, supervisors, or customers). These role-plays were intended to activate Agreeableness but were not identical to the situations in the SJT. Independent observers rated the degree to which participants expressed trait-related behaviors in the role-plays. They found that ITP scores for Agreeableness significantly correlated with self-reported Agreeableness ( $r$ s between .29 and .31; depending on ITPs as correlational or difference score) *and* significantly predicted Agreeableness scores in the role-plays ( $r$ s between .23 and .45) over and above self-reported Agreeableness.

Bledow and Frese (2009) developed an SJT for assessing personal initiative and investigated its construct-related and criterion-related validity of the SJT scores. They hypothesized that the assessment of personal initiative is compromised by Likert-type self-reports that ask for an abstract, generalized self-report because personal initiative is influenced by situational stimuli. Instead, Bledow and Frese (2009) argued that the situational character of personal initiative might be more appropriately measured via test-takers

preferences for personal initiative across different situations. Therefore, they developed an SJT that confronts test-takers with different, concrete job situations and response options which mirror actions that differ in their level of personal initiative. In line with their hypotheses, their SJT showed a positive, but only moderate relation to Likert-type self-reports of personal initiative. Moreover, personal initiative as assessed by the SJT predicted a unique amount of variance in supervisor ratings of personal initiative and overall performance that was not explained by Likert-type self-reports. SJTs and Likert-type self-reports thus appear to measure different parts of the construct domain of personal initiative. Or as another explanation the SJT prevented from socially desirable answers that might have been present in the Likert-type self-ratings. Furthermore, personal initiative as assessed by the SJT further mediated the effect of self-efficacy and felt responsibility on supervisor rated personal initiative and performance. Finally, personal initiative as assessed by the SJT converged with supervisor ratings of helping behavior, but not with supervisor ratings of conscientiousness. Thus, the pattern of convergent and discriminant relations indicates that the SJT measures test-takers' preferences for personal initiative across different situations on the job.

As a final example, Oostrom et al. (2019) developed an SJT that aims to measure the HEXACO personality dimensions. Following the guidelines from Guenole et al. (2017) and Lievens (2017b), they developed four items for each of the six HEXACO personality dimensions. Overall, Oostrom et al. (2019) found positive evidence for SJT scores' construct-related and criterion-related validity. For instance, SJT scores showed moderate to strong convergence with corresponding self- and other-reports. Further, SJT scores predicted criteria such as selection outcomes or supervisor rated organizational citizenship behavior. Honesty-Humility as assessed by the SJT was further found to predict supervisor rated organizational citizenship behavior after controlling for self-reports. In fact, Honesty-Humility as assessed

by the SJT mediated the effect of self-reported Honesty-Humility on organizational citizenship behavior.

In sum, according to emerging evidence, construct-driven SJTs can be seen as a promising alternative to common personality assessments. In fact, they show satisfactory psychometric properties (i.e., satisfactory factor structure, reliability, and convergent validity) and explain incremental variance over and above self-reported personality.

### **SJTs and the Personality-Situation Interplay**

In the previous sections, we discussed the traditional simulation-based view on SJTs and showed how more recent versions have started to measure constructs (personality traits). In this section, we go one step further and outline various strategies as to how SJTs might be further adapted to shed light on the personality-situation interplay.

#### **Assessment of Situation Construal**

As noted, SJTs consist of a situation component as well as a response component. In line with interactionism (Campion & Ployhart, 2013), it is assumed that candidates make sense of the short contextual SJT descriptions and that this situational construal guides their subsequent response choice. So, SJTs can be extended for use as a novel approach for assessing situation construal, which is of pivotal importance in promoting SJT use in the personality domain due to its role in many theories (e.g., CAPS, Mischel & Shoda, 1995). This might allow better explaining and understanding people's behavior, cognitions, and feelings.

One option to assess situation construal in SJTs was demonstrated by Rockstuhl, Ang, Ng, Lievens, and Van Dyne (2015). Their SJT did not present a list of predetermined (Multiple Choice) response options to individuals. Instead, they measured situation construal in an open-ended fashion by asking people to write down how they construed the situation. As another option, one might apply a multiple-choice approach in which one set of response

options depicts different alternatives on how to perceive the situation, whereas another set of response options describes different alternatives on how to react to the situation. Test-takers might then select one of the given situation construals that matches best their personal situation construal (see Lievens, 2017a). In this way, SJTs might serve as a viable tool to identify discrepancies between *individuals'* subjective construal and the *consensual* construal of a norm group (see also the section on Assessment of Personality Disorders below). As Rockstuhl (2017) noted, such insights might advance research on the subjective versus objective view of situations (see also Rauthmann, Sherman, & Funder, 2015).

Wood, Lowman, and Harms (2017) proposed to go one step further. They introduced the notion of field representations to more comprehensively capture the psychological situation. People's functional field representations of the situation refer to their mental simulations and expectancies of how various features of the situation covary with each other and with possible ways of acting in this particular situation. People's functional field representation are therefore supposed to subsequently influence their choice for or endorsement of different ways to respond to the situation. In SJTs, people's functional field representations might be assessed by not only asking people to submit the action they would choose but also to evaluate how their chosen action will affect several aspects of the situation., which can be valuable for determining whether people understand the more specific outcomes that should be pursued for making actions more or less effective responses (see Lowman, Wood, & Harms, 2017; Wood, Lowman, Harms, & Spain, 2019; Wood, Spain, & Harms, 2017).

### **Assessment of Within-Person Variability Across Situations**

Almost all SJTs compute people's mean score across the various items. However, SJTs and their inherent situation and response components lend themselves well to study meaningful variation in situation-trait contingencies because SJTs acknowledge that people

might adjust their response depending on the situation. SJTs thus enable computing two different sets of scores: an average score on each personality trait across the different SJT situations as well as a score reflecting variation in personality scores across these situations. Through the use of SJTs this means that researchers are able to assess both traditional cross-situational consistency (as reflected in mean trait level SJT scores) as well as within-person variability across situations (as reflected in standard deviation SJT scores associated with each trait).

Thus, SJTs that target specific traits also permit examining whether people vary their response indicative of a specific trait across situations (Dalal, Bhave, & Fiset, 2014). Recently, Lievens et al. (2018) found that people's within-person variability on sociability and dutifulness in responding across written SJT situations explained incremental variance over and above mean SJT scores and self-reports of functional flexibility in explaining peer ratings of performance in team projects. Importantly, the within-person variability as measured by the SJT was also related to actual personality state variability measured two years later via a 10-day experience-sampling study.

Viewed in light of the broader personality literature (Baird, Le, & Lucas, 2006; Erickson, Newman, & Pincus, 2009), computing an index of within-person variability associated with scores per construct across situations via SJTs is useful because the standardized SJT format enables to hold the situation constant. This is typically not the case when within-person variability is observed across situations/occasions via experience sampling methods (Fleeson & Gallagher, 2009; Furr, 2009).

In recent years, various analytical approaches have been used to go beyond analyses of SJT mean scores. For example, Lievens et al. (2018) measured within-person variability in responding across written SJT situations via IRT Tree Models (Böckenholt, 2012; De Boeck & Partchev, 2012; Lang, Lievens, De Fruyt, Zettler, & Tackett, 2019; Tutz, 1990). This IRT



approach enables distinguishing between (a) test-takers' mean-trait level/latent trait, (b) test-takers' trait variability as well as (c) items' mean-level/latent difficulty, and (d) items' propensity to elicit variability. To gain insight into the structure and response process underlying SJTs other studies applied other IRT models (e.g., Thiessen-Roe, 2013; Zu & Kyllonen, 2012) and generalizability theory (e.g., Jackson et al., 2017).

### **Assessment of Situation-Trait Contingencies**

Besides providing an index of within-person variability, one might also go one step further and examine how people's response option choices systematically vary according to specific situational dimensions or cues. This permits exploring whether there is meaningful group and individual variation in situation-trait contingencies (Huang & Ryan, 2011). Such "if-then" patterns are inherent parts of theoretical frameworks such as the Cognitive Affective Personality System (CAPS, Mischel & Shoda, 1995).

For both the assessment of within-person variability and situation-trait contingencies, it is important that the situations depicted in the SJT are carefully selected so that they mirror different types of psychological situations and trigger individuals to construe different situations in different ways. One strategy consists of developing the SJT situations on the basis of situational taxonomies (e.g., Parrigon et al., 2017; Rauthmann et al., 2014; Ziegler, 2014). For example, recently, Rauthmann et al. (2014) developed the DIAMONDS situational framework. These situational dimensions in this framework might be used to design a set of SJT item stems. In turn, SJT response options might reflect different levels of personality traits. In that case, it is possible to examine to what extent people's (between-person) or one person's (within-person) responses differ across and within situations that tap into a specific situational dimension. In addition, one might scrutinize how the inclusion of particular omnibus contextual elements (see the situation cues in the framework of Rauthmann et al., 2014) affects situation perception and SJT response option choice (see also

Brown, Jones, Serfass, & Sherman, 2016).

Such broad situational taxonomies (e.g., DIAMONDS, CAPTION) might predominantly capture situations in terms of their trait activation potential for specific traits. Therefore, as another strategy, the development of SJT situations might profit from more domain-specific situational taxonomies. For instance, to sample different leadership situations, test developers might rely on the taxonomy of different leadership situations described in the Multiple-Linkage Model (Yukl, 2010). To sample the interpersonal domain, test-developers might use the Interpersonal Circumplex Model (e.g., Carson, 1969; Kiesler, 1983). Clearly, future research related to broad and domain-specific situational taxonomies will benefit the development of SJTs to assess within-person variability and situation-trait contingencies.

### **Assessment of Proactive Transactions**

Traditionally, SJTs score low on interactivity because most SJTs are linear. That is, all applicants receive the same set of predetermined item situations and item options in the same sequence. So, the presentation of items is not dependent on people's responses to previous items. This linearity might be seen as contrived because it does not match how actual situations unfold. As the prototypical SJT puts constraints on individuals' propensities to select situations, shape existing situations, and create novel situations, it means that one only gets insight into personality reactivity (Judge, Hofmans, & Wille, 2017).

To avoid this drawback and shed light on personality-driven situation experience (Rauthmann, Sherman, Nave, & Funder, 2015), one might expand on the traditional SJT paradigm. That is, one might design SJTs so that an applicant's response to a situation determines the next situation that is presented. Accordingly, applicants are confronted with the consequences of their choices. This modality implies that all applicants do not respond to the same items. Such SJTs are called "branched", "nested", or "interactive" SJTs (Kanning et

al., 2006; Olson-Buchanan et al., 1998; see also Weekley et al., 2015). Conceptually, such non-linear/branched/game-like SJTs also enable going beyond the reactive transactions and digging deeper into proactive/manipulative/evocative transactions because people have the opportunity to enter, avoid, or shape specific situations (Baumert & Blum, 2017; see also Weekley et al., 2015).

For example, an SJT assessing extraversion might confront a test-taker with the situation of a new co-worker introducing himself to all coworkers in a large open office space. A possible response option indicative of higher levels of extraversion involves to actively start a conversation with the new colleague. Test-takers who endorse such a response option then receive an item that have them actively enter this conversation with the new colleague. In this new item, the new colleague mentions that he did not yet figure out where to get coffee. A possible response option reflective of higher levels of extraversion involves suggesting to have a short coffee break together, whereas a possible response option indicating lower levels of extraversion only mentions to the new colleague where to get coffee. In contrast, test-takers who did not endorse the response option to actively start a conversation with the new colleague would have avoided this situation. Instead, they might have endorsed a response option indicative of lower levels of extraversion (e.g., sending out an e-mail to the new colleague or not approaching the new colleague at all). These test-takers might then be confronted with a follow-up SJT item whose response options mirror different ways of answering or not answering an e-mail response from the new colleague. This example illustrates how interactive SJTs might serve to assess how test-takers actively enter, avoid, or shape situations. Accordingly, interactive SJTs also connect to the recent trend of gamification for assessment, selection, and development purposes (see, for example, Armstrong, Sanchez, & Bauer, 2017; Collmus & Landers, 2019; Georgiou, Gouras, & Nikolaou, 2019; Weekley et al., 2015; see also Miller, Jeong, & Christensen, this volume).

### **Assessment of Behavioral Responses**

Traditional SJTs do not require people to show behavior. SJTs are low-fidelity simulations of what people say they will do in different situations (instead of what they actually do). It is possible, though, to require respondents to enact their responses to a video-based situation in front of a webcam (e.g., Cucina, Su, Busciglio, Harris Thomas, & Thompson Peyton, 2015; Lievens, De Corte, & Westerveld, 2015; Lievens, Sackett, Dahlke, Oostrom, & De Soete, 2019). These webcam vignettes are then coded and rated by experienced assessors. Accordingly, a hybrid is created between SJTs and assessment center (i.e., low fidelity assessment center or high fidelity SJT). Interestingly, one might also compare the potential discrepancy between people's procedural knowledge (assessed via a written SJT) and their actually expressed behavior (assessed via assessment center exercises or via a webcam SJT; Breil, Geukes, & Back, 2017).

### **Assessment of Narratives and Goals**

Apart from assessing trait-based aspects of personality, there exist several options to widen the SJT paradigm to include the other two levels of personality: motivations/goals and narratives. For example, one might ask people to identify the story (narrative) that they associate with the scenarios. That is, people might be instructed to add short notes about past experiences that are similar to the scenarios depicted in the SJT (see Dunlop & Horton, 2017). Such elaboration might also be used to reduce faking good on SJTs (see Lievens & Peeters, 2008). In an extreme case, the SJT then becomes nothing more than a patterned behavior description interview, in which people are prompted to present a short narrative that provides information on their cognitions, emotions, and motivations for choosing specific behaviors in a specific situation and the goals they aimed to achieve. People's selection of the situations then also tells something about their personality.

Similarly, one might prompt people to write down the motivations/goals they have in the scenario. Although it is possible to design SJTs' response options so that they represent different strategies in goal achievement, these expansions of the SJT paradigm probably make most sense in an open-ended format (e.g., oral or written). For example, blank lines after the response options could request people to explain why they choose or endorse a specific response option and which specific motivation or goal this action might best serve. Such open-ended responses might provide insights into fundamental motivations such as needs for achievement, power, or affiliation (e.g., McClelland, 1987), regulatory focus or approach and avoidance motivation (e.g., Higgins, 1997; McClelland, 1987), and performance or learning goal orientation (e.g., Dweck, 1999).

### **Assessment of Personality Disorders**

So far, SJTs have mainly been used in employment settings. Yet, they might also be employed in clinical settings for assessing personality disorders. They then serve as an addition (or even alternative) to structured clinical interviews, self-report questionnaires, and ambulatory assessment approaches (Lievens, 2017; Wright & Simms, 2016). SJTs allow assessing how patients interpret and respond to a set of standardized scenarios and permit obtaining relevant diagnostic information. Examples are that they might indicate: endorsement of overtly problematic response options, too much rigidity in reacting to situations activating the same traits, extreme within-person variability (incoherence), and situation-trait contingencies that mismatch those of "normal" people. One might also add think-aloud procedures to SJT items to examine how people construe the situations (see section on Assessment of Situation Construal).

### **Epilogue**

So far, almost all SJT research has been conducted in employment and educational contexts. However, this chapter shows that SJTs also offer a lot of potential for the

personality psychology field and for shedding light on the person-situation interplay. The use of SJTs should also not be limited to assessing procedural knowledge about the costs and benefits of trait-related behavior in work-related contexts. Procedural knowledge about the effectiveness of trait-related courses of action is as important in other life domains such as relationship longevity, popularity among peers, emotion management, school achievement or health management. Moreover, this chapter illustrates that adaptations of SJTs might serve to assess people's situation construal, within-person variability across situations as well as situation-trait contingencies, proactive transactions, behavioral responses, narratives and goals as well as personality disorders. These broad application areas might further increase the applicability of SJTs in the broader field of psychology.

### References

- Armstrong, M. B., Sanchez, D. R., & Bauer, K. N. (2017, April). *Gaming and gamification IGNITE: Current trends in research and application*. Symposium conducted at the 32<sup>nd</sup> Annual Conference of the Society for Industrial and Organizational Psychology. Orlando, FL, USA.
- Arthur, W., Glaze, R. M., Jarrett, S. M., White, C. D., Schurig, I., & Taylor, J. E. (2014). Comparative evaluation of three situational judgment test response formats in terms of construct-related validity, subgroup differences, and susceptibility to response distortion. *Journal of Applied Psychology, 99*, 535–545.  
<https://doi.org/10.1037/a0035788>
- Baird, B. M., Le, K., & Lucas, R. E. (2006). On the nature of intraindividual personality variability: Reliability, validity, and associations with well-being. *Journal of Personality and Social Psychology, 90*, 512–527. <https://doi.org/10.1037/0022-3514.90.3.512>

- Baumert, A., & Blum, G. (2017). Employing situational simulations to understand person-situation transaction processes. *European Journal of Personality, 31*, 441–442.  
<https://doi.org/10.1002/per.2119>
- Bergman, M. E., Drasgow, F., Donovan, M. A., Henning, J. B., & Juraska, S. E. (2006). Scoring situational judgment tests: Once you get the data, your troubles begin. *International Journal of Selection and Assessment, 14*, 223–235.  
<https://doi.org/10.1111/j.1468-2389.2006.00345.x>
- Bledow, R., & Frese, M. (2009). A situational judgment test of personal initiative and its relationship to performance. *Personnel Psychology, 62*, 229–258.  
<https://doi.org/10.1111/j.1744-6570.2009.01137.x>
- Blickle, G., Momm, T. S., Kramer, J., Mierke, J., Liu, Y., & Ferris, G. R. (2009). Construct and criterion-related validation of a measure of emotional reasoning skills: A two-study investigation. *International Journal of Selection and Assessment, 17*, 101–118.  
<https://doi.org/10.1111/j.1468-2389.2009.00455.x>
- Böckenholt, U. (2012). Modeling multiple response processes in judgment and choice. *Psychological Methods, 17*, 665–678. <http://dx.doi.org/10.1037/a0028111>
- Breil, S. M., Geukes, K., & Back, M. D. (2017). Using situational judgment tests and assessment centres in personality psychology: Three suggestions. *European Journal of Personality, 31*, 442–443. <https://doi.org/10.1002/per.2119>
- Brown, N. A., Jones, A. B., Serfass, D. G., & Sherman, R. A. (2016). Reinvigorating the concept of a situation in situational judgment tests. *Industrial and Organizational Psychology, 9*, 38–42. <https://doi.org/10.1017/iop.2015.113>
- Bruce, M. M. (1974). *Examiner's manual: Supervisory Practices Test* (Rev. ed.). Larchmont, NY: Author.

- Callinan, M., & Robertson, I.T., 2000. Work sample testing. *International Journal of Selection and Assessment* 8, 248–260. <http://dx.doi.org/10.1111/1468-2389.00154>
- Campion, M. A., & Ployhart, R. E. (2013). Assessing personality with situational judgment measures: Interactionist psychology operationalized. In N. D. Christiansen & R. P. Tett (Eds.), *Handbook of personality at work* (pp. 439–456). New York, NY: Routledge.
- Cardall, A. J. (1942). *Preliminary manual for the test of practical judgment*. Chicago, IL: Science Research.
- Carson, R. C. (1969). *Interaction concepts of personality*. Chicago, IL: Aldine.
- Chan, D., & Schmitt, N. (2000). Interindividual differences in intraindividual changes in proactivity during organizational entry: A latent growth modeling approach to understanding newcomer adaptation. *Journal of Applied Psychology*, 85, 190–210. <https://doi.org/10.1037//0021-9010.85.2.190>
- Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology*, 63, 83–117. <https://doi.org/10.1111/j.1744-6570.2009.01163.x>
- Collmus, A. B., & Landers, R. N. (2019). Game-framing to improve applicant perceptions of cognitive assessments. *Journal of Personnel Psychology*, 18, 157-162. <https://doi.org/10.1027/1866-5888/a000227>
- Corstjens, J., Lievens, F., & Krumm, S. (2017). Situational judgment tests for selection: Distinguishing between contextualized and generic perspectives. In H. W. Goldstein, E. D. Pulakos, J. Passmore, & C. Semedo (Eds.), *The Wiley Blackwell Handbook of the psychology of recruitment, selection, and employee retention* (pp. 226–246). Hoboken, NJ: Wiley-Blackwell.



- Costello, C. K., Wood, D., & Tov, W. (2018). Revealed traits: A novel method for estimating cross-cultural similarities and differences in personality. *Journal of Cross-Cultural Psychology, 49*, 554-586. <https://doi.org/10.1177/0022022118757914>
- Cox, C. B., Barron, L. G., Davis, W., & de la Garza, B. (2017). Using situational judgment tests (SJTs) in training: Development and evaluation of a structured, low-fidelity scenario-based training method. *Personnel Review, 46*, 36-45. <https://doi.org/10.1108/PR-05-2015-0137>
- Cucina, J. M., Su, C., Busciglio, H. H., Harris Thomas, P., & Thompson Peyton, S. (2015). Video-based testing: A high-fidelity job simulation that demonstrates reliability, validity, and utility. *International Journal of Selection and Assessment, 23*, 197–209. <https://doi.org/10.1111/ijsa.12108>
- Dalal, R. S., Bhave, D. P., & Fiset, J. (2014). Within-person variability in job performance a theoretical review and research agenda. *Journal of Management, 40*, 1396–1436. <https://doi.org/10.1177/0149206314532691>
- De Boeck, P., & Partchev, I. (2012). IRTrees: Tree-based item response models of the GLMM family. *Journal of Statistical Software, 48*, 1–28. <http://dx.doi.org/10.18637/jss.v048.c01>
- De Soete, B. & Lievens, F. (2015). Situational judgment test. *International Encyclopedia of the Social & Behavioral Sciences, 22*, 13-19. <https://doi.org/10.1016/B978-0-08-097086-8.25092-7>
- Dunlop, W. L., & Horton, C. (2017). Two caveats to the adoption of personnel selection methods within personality psychology. *European Journal of Personality, 31*, 451–452. <https://doi.org/10.1002/per.2119>
- Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development*. Philadelphia, Psychology Press.

- Endler, N. S., Hunt, J. M., & Rosenstein, A. J. (1962). An S-R inventory of anxiousness. *Psychological Monographs: General and Applied*, 76, 1–33.  
<https://doi.org/10.1037/h0093817>
- Erickson, T. M., Newman, M. G., & Pincus, A. L. (2009). Predicting unpredictability: Do measures of interpersonal rigidity/flexibility and distress predict intraindividual variability in social perceptions and behavior? *Journal of Personality and Social Psychology*, 97, 893–912. <https://doi.org/10.1037/a0016515>
- File, Q. W. (1945). The measurement of supervisory quality in industry. *Journal of Applied Psychology*, 29, 323–337. <https://doi.org/10.1037/h0057397>
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51, 327–358.  
<https://doi.org/10.1037/h0061470>
- Fleeson, W., & Gallagher, P. (2009). The implications of Big Five standing for the distribution of trait manifestation in behavior: Fifteen experience-sampling studies and a meta-analysis. *Journal of Personality and Social Psychology*, 97, 1097–1114.  
<https://doi.org/10.1037/a0016786>
- Funke, U., & Schuler, H. (1998). Validity of stimulus and response components in a video test of social competence. *International Journal of Selection and Assessment*, 6, 115–123. <https://doi.org/10.1111/1468-2389.00080>
- Furnham, A., & Jaspars, J. (1983). The evidence for interactionism in psychology: A critical analysis of the situation-response inventories. *Personality and Individual Differences*, 4, 627–644. [https://doi.org/10.1016/0191-8869\(83\)90117-4](https://doi.org/10.1016/0191-8869(83)90117-4)
- Furr, M. R. (2009). Profile analysis in person–situation integration. *Journal of Research in Personality*, 43, 196–207. <https://doi.org/10.1016/j.jrp.2008.08.002>

- Georgiou, K., Gouras, A., & Nikolaou, I. (2019). Gamification in employee selection: The development of a gamified assessment. *International Journal of Selection and Assessment*, 27, 91-103. <https://doi.org/10.1111/ijsa.12240>
- Gerpott, F. H., Thielmann, I., & Balliet, D. (under review). Interdependence approaches to the person and situation. In D. Wood, S. J. Read, P. D. Harms, & A. Slaughter (Eds.): *Emerging approaches to measuring and modeling the person and situation* (Chapter 17). Amsterdam, NL: Elsevier.
- Greenberg, S. H. (1963). *Supervisory judgment test manual*. Washington, DC: US Civil Service Commission.
- Guenole, N., Chernyshenko, O. S., & Weekly, J. (2017). On designing construct driven situational judgment tests: Some preliminary recommendations. *International Journal of Testing*, 17, 234–252. <https://doi.org/10.1080/15305058.2017.1297817>
- Hauenstein, N. M., Findlay, R. A., & McDonald, D. P. (2010). Using situational judgment tests to assess training effectiveness: Lessons learned evaluating military equal opportunity advisor trainees. *Military Psychology*, 22, 262-281. <https://doi.org/10.1080/08995605.2010.492679>
- Herde, C. N., Lievens, F., Jackson, D. J. R., Shalfrooshan, A., & Roth, P. L. (in press). Subgroup differences in situational judgment test scores: Evidence from large applicant samples. *International Journal of Selection and Assessment*.
- Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist*, 52, 1280-1300. <http://dx.doi.org/10.1037/0003-066X.52.12.1280>
- Hooper, A. C., Cullen, M. J., & Sackett, P. R. (2006). Operational threats to the use of SJTs: Faking, coaching and retesting issues. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests. Theory, measurement, and application* (pp. 205–232). Mahwah, NJ: Erlbaum.

- Huang, J. L., & Ryan, A. M. (2011). Beyond personality traits: A study of personality states and situational contingencies in customer service jobs. *Personnel Psychology, 64*, 451–488. <https://doi.org/10.1111/j.1744-6570.2011.01216.x>
- Hunter, D. R. (2003). Measuring general aviation pilot judgment using a situational judgment technique. *The International Journal of Aviation Psychology, 13*, 373–386. [https://doi.org/10.1207/S15327108IJAP1304\\_03](https://doi.org/10.1207/S15327108IJAP1304_03)
- Jackson, D. J. R., LoPilato, A. C., Hughes, D., Guenole, N., & Shalfrooshan, A. (2017). The internal structure of situational judgement tests reflects candidate main effects: Not dimensions or situations. *Journal of Occupational and Organizational Psychology, 90*, 1-27. <https://doi.org/10.1111/joop.12151>
- Judge, T. A., Hofmans, J., & Wille, B. (2017). Situational judgement tests and personality measurement: Some answers and more questions. *European Journal of Personality, 31*, 463–464. <https://doi.org/10.1002/per.2119>
- Kanning, U. P., Grewe, K., Hollenberg, S., & Hadouch, M. (2006). From the subjects' point of view. *European Journal of Psychological Assessment, 22*, 168–176. <https://doi.org/10.1027/1015-5759.22.3.168>
- Kasten, N., Freund, P. A., & Staufenbiel, T. (2018). Sweet little lies. An in-depth analysis of faking behavior on situational judgment tests compared to personality questionnaires. *European Journal of Psychological Assessment*. <https://doi.org/10.1027/1015-5759/a000479>
- Kiesler, D. J. (1983). The 1982 interpersonal circle: A taxonomy for complementarity in human transactions. *Psychological Review, 90*, 185–214. <http://dx.doi.org/10.1037/0033-295X.90.3.185>

- Konradt, U., Hertel, G., & Joder, K. (2003). Web-based assessment of call center agents: Development and validation of a computerized instrument. *International Journal of Selection and Assessment, 11*, 184–193. <https://doi.org/10.1111/1468-2389.00242>
- Krumm, S., Lievens, F., Hüffmeier, J., Lipnevich, A. A., Bendels, H., & Hertel, G. (2015). How “situational” is judgment in situational judgment tests? *Journal of Applied Psychology, 100*, 399–416. <https://doi.org/10.1037/a0037674>
- Lang, J. W., Lievens, F., De Fruyt, F., Zettler, I., & Tackett, J. L. (2019). Assessing meaningful within-person variability in Likert-scale rated personality descriptions: An IRT tree approach. *Psychological Assessment, 31*, 474–487. <http://dx.doi.org/10.1037/pas0000600>
- Lievens, F. (2017a). Assessing personality–situation interplay in personnel selection: Toward more integration into personality research. *European Journal of Personality, 31*, 424–440. <https://doi.org/10.1002/per.2111>
- Lievens, F. (2017b). Construct-driven SJTs: Toward an agenda for future research. *International Journal of Testing, 17*, 269–276. <https://doi.org/10.1080/15305058.2017.1309857>
- Lievens, F., De Corte, W., & Westerveld, L. (2015). Understanding the building blocks of selection procedures: Effects of response fidelity on performance and validity. *Journal of Management, 41*, 1604–1627. <https://doi.org/10.1177/0149206312463941>
- Lievens, F., & De Soete, B. (2012). Simulations. In N. Schmitt (Ed.), *The Oxford handbook of personnel assessment and selection* (pp. 383–410). Oxford, UK: Oxford University Press.
- Lievens, F., Lang, J. W. B., De Fruyt, F., Corstjens, J., Van de Vijver, M., & Bledow, R. (2018). The predictive power of people’s intraindividual variability across situations:

- Implementing whole trait theory in assessment. *Journal of Applied Psychology, 103*, 753–771. <https://doi.org/10.1037/apl0000280>
- Lievens, F., & Motowidlo, S. J. (2016). Situational judgment tests: From measures of situational judgment to measures of general domain knowledge. *Industrial and Organizational Psychology, 9*, 3–22. <https://doi.org/10.1017/iop.2015.71>
- Lievens, F., & Patterson, F. (2011). The validity and incremental validity of knowledge tests, low-fidelity simulations, and high-fidelity simulations for predicting job performance in advanced-level high-stakes selection. *Journal of Applied Psychology, 96*, 927–940. <http://dx.doi.org/10.1037/a0023496>
- Lievens, F., & Peeters, H. (2008). Impact of elaboration on responding to situational judgment test items. *International Journal of Selection and Assessment, 16*, 345–355. <https://doi.org/10.1111/j.1468-2389.2008.00440.x>
- Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: A review of recent research. *Personnel Review, 37*, 426–441. <https://doi.org/10.1108/00483480810877598>
- Lievens, F., & Sackett, P. R. (2012). The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *Journal of Applied Psychology, 97*, 460–468. <https://doi.org/10.1037/a0025741>
- 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. <http://dx.doi.org/10.1037/a0014628>
- Lievens, F., Sackett, P. R., Dahlke, J. A., Oostrom, J. K., & De Soete, B. (2019). Constructed response formats and their effects on minority–majority differences and validity. *Journal of Applied Psychology, 104*, 715–726. <http://dx.doi.org/10.1037/apl0000367>

- Lowman, G., Wood, D., & Harms, P. D. (2017). *Assertiveness and power in the workplace: A functional field approach*. University of Alabama, Tuscaloosa, AL.
- MacCann, C., & Roberts, R. D. (2008). New paradigms for assessing emotional intelligence: Theory and data. *Emotion, 8*, 540–551. <https://doi.org/10.1037/a0012746>
- McClelland, D. C. (1987). *Human motivation*. Cambridge: Cambridge University Press.
- McClough, A. C., & Rogelberg, S. G. (2003). Selection in Teams: An exploration of the teamwork knowledge, skills, and ability test. *International Journal of Selection and Assessment, 11*, 56–66. <https://doi.org/10.1111/1468-2389.00226>
- McCrae, R. R., & Costa, P., Jr. (1996). Toward a new generation of personality theories: Theoretical contexts for the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 51–87). New York, NY: Guilford.
- 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, 86*, 730-740. <https://doi.org/10.1037/0021-9010.86.4.730>
- McDaniel, M. A., Hartman, N., Whetzel, D. L., & Grubb III, W. L. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology, 60*, 63–91. <https://doi.org/10.1111/j.1744-6570.2007.00065.x>
- McHenry, J. J., & Schmitt, N. (1994). Multimedia testing. In M. G. Rumsey & C. B. Walker (Eds.), *Personnel selection and classification* (pp. 193–232). Hillsdale, NJ: Erlbaum.
- Miller, L., Jeong, D., & Christensen, J. Game-based testing (under review). Interdependence approaches to the person and situation. In D. Wood, S. J. Read, P. D. Harms, & A. Slaughter (Eds.), *Emerging approaches to measuring and modeling the person and situation*. Amsterdam, NL: Elsevier.

- Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, *102*, 246–268. <https://doi.org/10.1037/0033-295X.102.2.246>
- Morgeson, F. P., Reider, M. H., & Campion, M. A. (2005). Selecting individuals in team settings: The importance of social skills, personality characteristics, and teamwork knowledge. *Personnel Psychology*, *58*, 583–611. <https://doi.org/10.1111/j.1744-6570.2005.655.x>
- Motowidlo, S. J., & Beier, M. E. (2010). Differentiating specific job knowledge from implicit trait policies in procedural knowledge measured by a situational judgment test. *Journal of Applied Psychology*, *95*, 321–333. <https://doi.org/10.1037/a0017975>
- 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. <http://dx.doi.org/10.1037/0021-9010.75.6.640>
- Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006). Implicit policies about relations between personality traits and behavioral effectiveness in situational judgment items. *Journal of Applied Psychology*, *91*, 749–761. <https://doi.org/10.1037/0021-9010.91.4.749>
- Mussel, P., Gatzka, T., & Hewig, J. (2018). Situational judgment tests as an alternative measure for personality assessment. *European Journal of Psychological Assessment*, *34*, 328–335. <https://doi.org/10.1027/1015-5759/a000346>
- Northrop, L. C. (1989). *The psychometric history of selected ability constructs*. Washington, DC: United State Office of Personnel Management.
- Olson-Buchanan, J. B., & Drasgow, F. (2006). Multimedia situational judgment tests: The medium creates the message. In *SIOP organizational series. Situational judgment*



*tests: Theory, measurement, and application* (pp. 253–278). Mahwah, NJ, US:

Lawrence Erlbaum Associates Publishers.

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. <https://doi.org/10.1111/j.1744-6570.1998.tb00714.x>

Oostrom, J. K., de Vries, R. E., & de Wit, M. (2019). Development and validation of a

HEXACO situational judgment test. *Human Performance*, *32*, 1–29.

<https://doi.org/10.1080/08959285.2018.1539856>

Ostroff, C. (1991). Training effectiveness measures and scoring schemes: A comparison.

*Personnel Psychology*, *44*, 353–374. <https://doi.org/10.1111/j.1744->

[6570.1991.tb00963.x](https://doi.org/10.1111/j.1744-6570.1991.tb00963.x)

Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing

a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology*, *89*, 187–207.

<https://doi.org/10.1037/0021-9010.89.2.187>

Parrigon, S., Woo, S. E., Tay, L., & Wang, T. (2017). CAPTION-ing the situation: A

lexically-derived taxonomy of psychological situation characteristics. *Journal of*

*Personality and Social Psychology*, *112*, 642. <https://doi.org/10.1037/pspp0000111>

Patterson, F., Zibarras, L., & Ashworth, V. (2016). Situational judgement tests in medical

education and training: Research, theory and practice: AMEE Guide No. 100.

*Medical Teacher*, *38*, 3–17. <https://doi.org/10.3109/0142159X.2015.1072619>

Peeters, H., & Lievens, F. (2005). Situational judgment tests and their predictiveness of

college students' success: The influence of faking. *Educational and Psychological*

*Measurement*, *65*, 70–89. <https://doi.org/10.1177/0013164404268672>

- Ployhart, R. E., & MacKenzie Jr., W. I. (2011). Situational judgment tests: A critical review and agenda for the future. In *APA Handbooks in Psychology. APA handbook of industrial and organizational psychology, Vol 2: Selecting and developing members for the organization* (pp. 237–252). <https://doi.org/10.1037/12170-008>
- Prasad, J. J., Showler, M. B., Schmitt, N., Ryan, A. M., & Nye, C. D. (2017). Using biodata and situational judgment inventories across cultural groups. *International Journal of Testing, 17*, 210–233. <https://doi.org/10.1080/15305058.2016.1218338>
- Rauthmann, J., Gallardo-Pujol, D., Guillaume, E., Todd, E., Nave, C., Sherman, R. A., ... Funder, D. C. (2014). The situational eight DIAMONDS: A taxonomy of major dimensions of situation characteristics. *Personality Processes and Individual Differences, 107*, 677–718. <http://dx.doi.org/10.1037/a0037250>
- Rauthmann, J. F., Sherman, R. A., & Funder, D. C. (2015). Principles of situation research: Towards a better understanding of psychological situations. *European Journal of Personality, 29*, 363-381. <https://doi.org/10.1002/per.1994>
- Rauthmann, J. F., Sherman, R. A., Nave, C. S., & Funder, D. C. (2015). Personality-driven situation experience, contact, and construal: How people's personality traits predict characteristics of their situations in daily life. *Journal of Research in Personality, 55*, 98–111. <https://doi.org/10.1016/j.jrp.2015.02.003>
- Rockstuhl, T. (2017). Situational construals: Key to assessments of personality-situation interplay. *European Journal of Personality, 31*, 481-482. <https://doi.org/10.1002/per.2119>
- Rockstuhl, T., Ang, S., Ng, K.-Y., Lievens, F., & Van Dyne, L. (2015). Putting judging situations into situational judgment tests: Evidence from intercultural multimedia SJTs. *Journal of Applied Psychology, 100*, 464–480. <https://doi.org/10.1037/a0038098>

- Roth, P. L., Bobko, P., & Buster, M., A. (2013). Situational judgment tests: The influence and importance of applicant status and targeted constructs on estimates of Black–White subgroup differences. *Journal of Occupational and Organizational Psychology*, *86*, 394–409. <https://doi.org/10.1111/joop.12013>
- Schäpers, P., Lievens, F., Freudenstein, J.-P., Hüffmeier, J., König, C. J., & Krumm, S. (in press). Removing situation descriptions from situational judgment test items: Does the impact differ for video-based versus text-based formats? *Journal of Occupational and Organizational Psychology*. <https://doi.org/10.1111/joop.12297>
- Schäpers, P., Mussel, P., Lievens, F., König, C. J., Freudenstein, J.-P., & Krumm, S. (in press). The role of situations in situational judgment tests: Effects on construct saturation, predictive validity, and applicant perceptions. *Journal of Applied Psychology*. <https://doi.org/10.1037/apl0000457>
- Schmitt, N., & Ostroff, C. (1986). Operationalizing the “behavioral consistency” approach: Selection test development based on a content-oriented strategy. *Personnel Psychology*, *39*, 91–108. <https://doi.org/10.1111/j.1744-6570.1986.tb00576.x>
- Stevens, M. J., & Campion, M. A. (1999). Staffing work teams: Development and validation of a selection test for teamwork settings. *Journal of Management*, *25*, 207–228. [https://doi.org/10.1016/S0149-2063\(99\)80010-5](https://doi.org/10.1016/S0149-2063(99)80010-5)
- Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology*, *88*, 500–517. <https://doi.org/10.1037/0021-9010.88.3.500>
- Tett, R. P., & Guterman, H. A. (2000). Situation trait relevance, trait expression, and cross-situational consistency: Testing a principle of trait activation. *Journal of Research in Personality*, *34*, 397–423. <https://doi.org/10.1006/jrpe.2000.2292>

- Thissen-Roe, A. (2013). Modeling situational judgment items with multiple distractor dimensions. In R. E. Millsap, L. A. van der Ark, D. M. Bolt, & C. M. Woods (Eds.), *New Developments in Quantitative Psychology* (pp. 251-265). Springer, New York, NY.
- Thornton III, G. C., Mueller-Hanson, R. A., & Rupp, D. E. (2017). *Developing organizational simulations. A guide for practitioners, students, and researchers*. New York, NY: Routledge. <https://doi.org/10.4324/9781315652382>
- Tutz, G. (1990). Sequential item response models with an ordered response. *British Journal of Mathematical & Statistical Psychology*, *43*, 39–55.  
<http://dx.doi.org/10.1111/j.2044-8317.1990.tb00925.x>
- 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. <https://doi.org/10.1146/annurev-orgpsych-032414-111304>
- Weekley, J. A., Ployhart, R. E., & Holtz, B. C. (2006). On the development of situational judgment tests: Issues in item development, scaling, and scoring. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 157-182). Mahwah, NJ: Erlbaum.
- Wernimont, P. F., & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology*, *52*, 372. <http://dx.doi.org/10.1037/h0026244>
- Westring, A. J. F., Oswald, F. L., Schmitt, N., Drzakowski, S., Imus, A., Kim, B., & Shivpuri, S. (2009). Estimating trait and situational variance in a situational judgment test. *Human Performance*, *22*, 44–63. <https://doi.org/10.1080/08959280802540999>
- Whetzel, D. L., & McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review*, *19*, 188-202.  
<https://doi.org/10.1016/j.hrmr.2009.03.007>

- Wood, D., Lowman, G. H., & Harms, P. D. (2017). 'Low-fidelity simulations' play central roles in explaining behaviour. *European Journal of Personality*, *31*, 483–484.  
<https://doi.org/10.1002/per.2119>
- Wood, D., Lowman, G. H., Harms, P. D., & Spain, S. M. (2019). Using functional fields to formally represent the meaning and logic of behavior: A worked example using Dark Triad-related actions. *Personality and Individual Differences*, *136*, 24-37.  
<https://doi.org/10.1016/j.paid.2017.09.002>
- Wood, D., Spain, S. M., & Harms, P. D. (2017). Functional approaches to representing the interplay of situations, traits, and behavior. In D. C. Funder, R. A. Sherman, & J. F. Rauthmann (Eds.), *Handbook of psychological situations*. New York: Oxford.
- Wood, D., Tov, W., & Costello, C. (2015). What a\_ thing to do! Formally characterizing actions by their expected effects. *Journal of Personality and Social Psychology*, *108*, 953-976. <http://dx.doi.org/10.1037/pspp0000030>
- Wright, A. G. C., & Simms, L. J. (2016). Stability and fluctuation of personality disorder features in daily life. *Journal of Abnormal Psychology*, *125*, 641–656.  
<http://dx.doi.org/10.1037/abn0000169>
- Yukl, G. (2010). *Leadership in Organizations*. Upper Saddle River, NJ: Prentice Hall.
- Ziegler, M. (2014). B5PS. Big Five Inventory of personality in occupational situations. Mödling, Austria: Schuhfried GmbH.
- Zu J., & Kyllonen P. C. 2012. *Scoring situational judgment tests with item response models*. Rep. ETS-2012-0160, R1, Educational Testing Service: Princeton, NJ.

Table 1

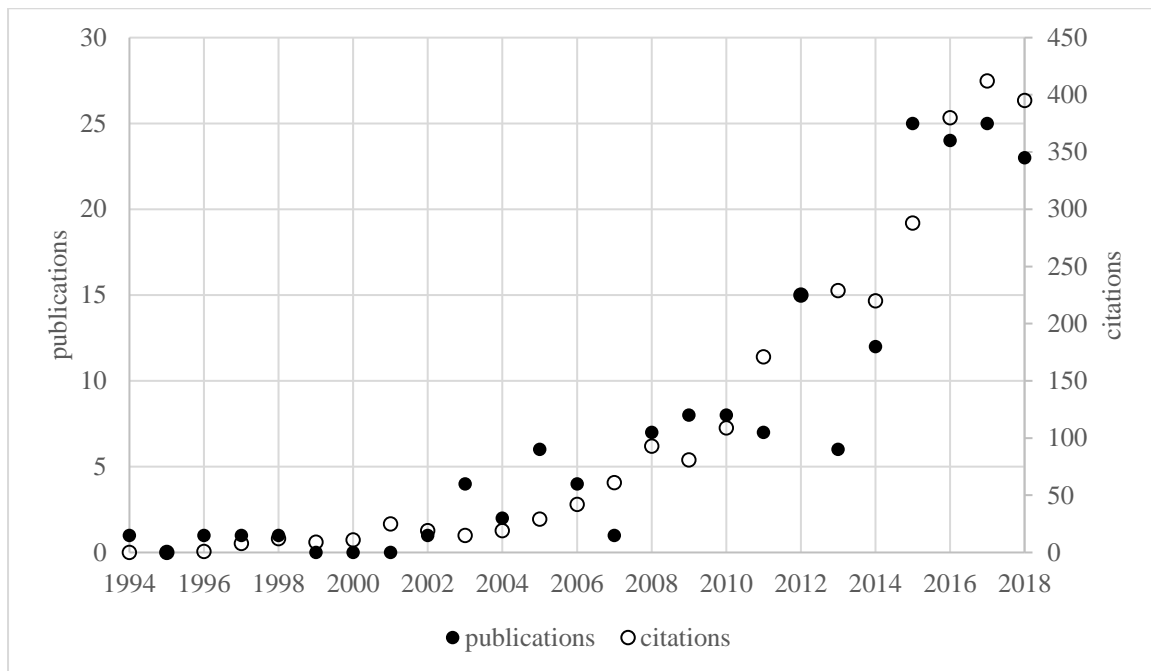
*Differences in Development of Traditional versus Construct-Driven SJTs.*

	Traditional SJTs	Construct-driven SJTs
Dimensionality	Multidimensionality (heterogeneity) at the item level	Unidimensionality (homogeneity) at the item level
Internal consistency	Usually low to moderate internal consistency (for longer SJTs)	Satisfactory internal consistency
Development of situations	Critical incident technique with input from job holders and SMEs	Input from psychologists and psychological theories (e.g., trait activation theory and situational taxonomies)
Development of response options	Job incumbents generate response options that might tap various constructs. Effectiveness-ratings by SMEs and psychologists.	Input from psychologists to develop item responses reflect different levels of the construct of interest.
Response instructions	Knowledge (“Should do”) and behavioral (“Would do”) instructions	Knowledge (“Should do”) and behavioral (“Would do”) instructions
Scoring strategy	Effectiveness scoring	Effectiveness and trait scoring

*Note.* See (Lievens, 2017b).

Figure 1.

*Publications on SJTs and citations from 1994-2018.*



*Note.* This figure was created based upon a Web of Science search for “situational judgment test(s)” (31/12/2018).