

## Assessing Personality–Situation Interplay in Personnel Selection: Toward More Integration into Personality Research

FILIP LIEVENS\*

*Department of Personnel Management and Work and Organizational Psychology, Ghent University, Ghent, Belgium*

*Abstract: Over the years, the personnel selection field has developed methods to assess trait expression in particular situations, but these approaches have evolved mostly outside the field of personality psychology. In this article, I review available personnel selection evidence regarding two such approaches: (i) situational judgement tests that present short scenarios and ask job candidates how they would handle the situations and (ii) assessment centre exercises requiring candidates to display behaviour in specified interactive situations. I describe these approaches and discuss their relations with personality research. I posit that adapting these approaches to personality research creates methodological diversity to address key research themes related to within-person variability, trait–behaviour links, personality disorders, and personality expression and perception. Copyright © 2017 European Association of Personality Psychology*

Contemporary conceptualizations of personality have expanded traditional definitions of personality beyond purely dispositional models to include situational/contextual components and intraindividual variability in responding to these situations. For example, personality ‘states’ refer to momentary displays of personality traits on particular occasions (Fleeson & Gallagher, 2009). In addition, behavioural ‘signatures’ of personality denote meaningful within-individual patterns of situation–behaviour relations (Mischel, 2009; Mischel & Shoda, 1995). According to these contemporary conceptualizations, people’s variability across occasions/situations is viewed as potentially substantively meaningful (Moskowitz & Zuroff, 2004). There is emerging consensus that both consistency and within-person variability across situations should take important places in personality conceptualizations (e.g. Dunlop, 2015; Fleeson, 2001; Hogan & Roberts, 2000; Johnson, 2015; Judge, Simon, Hurst, & Kelley, 2014; Roberts, 2007), as reflected in, for example, ‘whole trait theory’ (Fleeson & Jayawickreme, 2015).

Concordant with these conceptual personality developments, a variety of measurement methods other than traditional self-reports have been proposed and used. Recently, Wrzus and Mehl (2015) reviewed such lab and field approaches (e.g. experience-sampling methods and behaviour observation). Parallel to these conceptual and methodological developments in the personality domain, personnel selection researchers have also developed and examined personality measurement approaches focused on trait expression in particular situations. In particular, there

exist rich research bases about two such personnel selection approaches (situational judgement tests and assessment centre exercises)<sup>1</sup> that focus on the personality–situation interplay, but these approaches have evolved largely outside the field of personality.

Bringing these two approaches to the attention of personality researchers is timely for two reasons. First, it fits well with contemporary personality conceptualizations emphasizing within-person variability across situations and situation–trait contingencies. Second, in current personality research, most measurement approaches to within-person variability (e.g. experience-sampling) have not assessed the specific situations eliciting the behaviours (Geukes, Nestler, Hutteman, Küfner, & Back, in press). More generally, Wrzus and Mehl (2015) noted that ‘it is the measurement of situations that historically has been the sore spot of personality (and social) psychology..., and it is there where novel measurement approaches may have the biggest leverage’ (p. 265). Given this current need for expanding personality measurement, this article (i) reviews the strands of research on situational judgement tests and assessment centre exercises and (ii) describes how these two approaches allow tackling key questions in personality research. Accordingly, I hope to start a constructive dialogue between selection and personality researchers to help both research areas enrich their theoretical and empirical approaches.<sup>2</sup>

\*Correspondence to: Filip Lievens, Department of Personnel Management and Work and Organizational Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium.  
E-mail: filip.lievens@ugent.be

Previous versions of this paper were presented as invited keynote at the European Congress of Personality (2014, Lausanne, Switzerland) and at the American Psychological Society (2016, Chicago, IL, USA).

<sup>1</sup>Selection researchers have also explored other approaches, including contextualized personality inventories (Shaffer & Postlethwaite, 2012), other reports (Connelly & Ones, 2010; Oh, Wang, & Mount, 2011), implicit personality measures (James, 1998), structured interviews (Van Iddekinge, Raymark, & Roth, 2005), and resume analysis (Ragsdale, Rahael, Christiansen, Frost, & Burns, 2013).

<sup>2</sup>As personnel selection research has drawn on many developments in personality psychology, this article mainly focuses on how personality research can benefit from selection research.

*Handling editor: Wendy Johnson*

I begin by outlining the characteristics and goals of selection research (as compared with personality research) and the reasons behind selection researchers' exploration of alternative personality assessments. In the next two main sections, I review the research streams on situational judgement tests and assessment centres, respectively. Per research stream, I list the key elements of the selection procedure, summarize relevant research evidence, zoom into the differences and similarities with existing approaches in personality research, and show how the procedure can be integrated into the personality psychology field to address relevant questions in personality psychology. I end by giving general directions for future research and/or concrete examples of potential studies in the personality domain.

## PERSONNEL SELECTION: GOALS AND RATIONALES

### Goal and focus of personnel selection

Selection settings are characterized by high stakes for both candidates (whose selection decisions impact their careers) and employers (who need employees who will perform well on the job). Generally, the goal of personnel selection is to identify whether candidates have the knowledge, skills, abilities, and other characteristics (KSAOs) needed to perform effectively in particular jobs (Gatewood, Field, & Barrick, 2015). Personality traits are among these KSAOs. To reach their goal, personnel selectors use prespecified procedures to assess whether a focal person has the required KSAOs and to assess whether this person's standings on these KSAOs will be satisfactory to perform the job. This goal creates an important difference from personality research. Due to the job-context focus of selection research, effectiveness of behaviour is of primary relevance, and this effectiveness is assessed relative to criteria specified not by the focal person, but by another party (the employer). Personality is related to effectiveness, but only moderately at best. In personality research, however, the focus of interest is on personality itself, typically on how and why personality is structured, developed, and manifested throughout life, whether 'effective' or not. Where satisfaction with behaviour/role/context is considered, it is generally from the perspective of the focal person.

There are also points of correspondence between these two domains. First, personality research has also examined situations outside the job context in which others' approval is directly relevant to varying degrees, such as making friends in new environments, resolving conflicts, and romantic dating. More generally, research on reputation (Hofstee, 1994; Hogan & Shelton, 1998; Vazire, 2010) and social effects of personality (e.g. popularity in a group; Asendorpf & Wilpers, 1998; Back et al., 2011; Back & Vazire, 2015; Leckelt, Küfner, Nestler, & Back, 2015; Mehl, Gosling, & Pennebaker, 2006) are areas of interest of personality researchers. Second, people manage the impressions they make (as is common in selection settings due to the high stakes) in many social (evaluative) interactions and life domains studied in personality research (Levy, Collins, &

Nail, 1998) and how and to what extent they do is even considered an area of interest itself (e.g. Giddens, 1991; Goffman, 1959; Schlenker & Leary, 1982; Snyder, 1974). Finally, in this article, I also suggest ways that personality researchers can tailor the approaches discussed to their needs (e.g. by placing the example situations in a wider variety of contexts), and I suspect that personality researchers can come up with even more creative ways to do so.

### Rationales for the search for other personality measurement approaches in selection

#### *Response distortion in self-report personality inventories*

Self-reports of personality are vulnerable to intentional response distortion ('faking good'). For decades, this issue has been hanging like a sword of Damocles over the utility of generalized self-report personality inventories in personnel selection (Sackett, Lievens, Van Iddekinge, & Kuncel, 2017). Such distortion has been shown to result in higher scores on job-relevant traits, lower standard deviations of scores, and higher intertrait correlations. Although effects of faking good on criterion-related validities of personality scores are more mixed, the overall sentiment is that self-report personality measures should show better criterion-related validity than they have (i.e., the highest validity coefficient is 0.20 for conscientiousness: Hurtz & Donovan, 2000; see also Barrick & Mount, 1991; Sackett & Lievens, 2008). Perhaps more important, fakers tend to rise to the tops of the score distributions, making likely better, honest candidates less likely to be selected (e.g. Mueller-Hanson, Heggstad, & Thornton, 2003). Therefore, alternatives to traditional generalized self-report personality inventories that are less prone to faking good have been sought (Connelly & Ones, 2010).

#### *Incorporating work situations into personality assessment*

Selection researchers have argued that incorporating work situations (or descriptions thereof) directly into personality assessment can improve ability to hire the best-suited candidates over reliance on generalized personality self-report measures (Campion & Ployhart, 2013). Therefore, selection researchers have explored more contextualized forms of personality assessment. Contextualized personality inventories that add situational keywords (e.g. 'I pay attention to details *at work*') and specify to candidates which context (situational referent) they should consider when responding to the statements exemplify how increased contextualization might improve prediction (Lievens, De Corte, & Schollaert, 2008; Shaffer & Postlethwaite, 2012; see also Bleidorn & Ködding, 2013). In fact, the most recent meta-analysis of Shaffer and Postlethwaite showed that for four Big Five traits (emotional stability, extraversion, agreeableness, and openness to experience) the sizes of validity coefficients of ratings on contextualized personality inventories were at least double those of generalized inventories for predicting the same broad criterion of job performance, consistent with analogous observations for similarly matched contextualization and outcomes in personality research (e.g. Clifton, 2014; Slatcher & Vazire,

2009). As will be argued in the succeeding texts, the two approaches of interest in this article go even further in embedding contextualized stimuli in assessment.

#### *Candidates' perceptions of self-report personality inventories*

A third reason for searching for alternative personality assessment methodologies stems from observations that candidates' perceptions of selection procedures matter (McCarthy et al., 2013). How candidates view selection procedures affects their test motivation/test performance and, in some cases, the validity of the selection procedure. Moreover, candidates use selection procedures as signals to make inferences about employer culture (Bangerter, Roulin, & König, 2012). Especially when there is high demand for specific and uncommon candidate profiles, it is important to employers that qualified candidates have favourable perceptions of the selection procedures and the hiring processes. Therefore, candidates' reactions to selection procedures have generated a prolific line of research. The most recent meta-analysis (Hausknecht, Day, & Thomas, 2004) revealed that self-report personality inventories received an average favourability rating of 2.88 out of 5 (see their Table 4, p. 659). This ranked personality inventories sixth among the 10 selection procedures examined (employment interviews and work samples received the highest ratings). Common reasons put forward by candidates for this relatively low rating include perceived low job relatedness, rather impersonal nature (compared, for instance, to employment interviews), and lack of opportunity to demonstrate competence (compared, for example, with work samples). Given lower favourability ratings of personality inventories and the reasons for them, selection researchers have explored alternatives that might be more favourably perceived by job candidates. The two approaches discussed in the succeeding texts fulfil this criterion.

## SITUATIONAL JUDGEMENT TESTS

### **SJTs: Definition, characteristics, and scoring**

Over the years, selection researchers have invested substantial effort in development and use of Situational Judgement Tests (SJTs). In SJTs, candidates are presented with short job-related situational descriptions and various response options to handle these situations (McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001). SJTs tend to be favourably perceived by candidates (Kanning, Grewe, Hollenberg, & Hadouch, 2006).

Situational Judgement Tests are not new inventions; early versions appeared before WWII. In recent years, though, there has been growing interest in developing SJTs specifically to assess personality traits that are considered job relevant (e.g. Bledow & Frese, 2009; Campion & Ployhart, 2013; Christian, Edwards, & Bradley, 2010; Labrador, Christiansen, & Burns, 2006; Lievens & Motowidlo, 2015; Motowidlo, Hooper, & Jackson, 2006a, 2006b; Mussel, Gatzka, & Hewig, 2016). SJT items present job-related situations that are relevant to

traits that matter in the targeted jobs (e.g. extraversion for sales-related jobs). An SJT item situation typically provides brief information about what is going on or at issue in the situation, the people involved in it (e.g. occupation and gender), the situation's relative novelty (e.g. is this the first time it happens?), and the place (e.g. office). In most SJT items, the situations are still presented in text-based formats, although more life-like formats (e.g. multimedia, 2D animation, 3D animation, and even avatars) are increasingly used. After presenting the SJT situation, candidates are asked 'What would you do?' (i.e., an instruction to indicate behavioural preference).<sup>3</sup> A list of response options follows. Various response formats exist: Candidates might be asked to pick one response option from the list, rank the options ('What would you prefer doing most, least?') or rate them ('How strongly does each option reflect what you would do?'). As SJTs typically ask people what they would do instead of showing actual behaviour, they are also referred to as low-fidelity simulations and tacit knowledge inventories.

When developing SJTs that target personality traits, the response options are written to represent different levels of the targeted traits. To generate the response options, SJT developers typically rely on groups of subject-matter experts (e.g. graduate students familiar with personality trait psychology) 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, e.g. Motowidlo et al., 2006a; Mussel et al., 2016). 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 scoring 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 (Bergman, Drasgow, Donovan, Henning, & Juraska, 2006). Table 1 summarizes the characteristics and scoring of SJTs and compares them with traditional self-report personality inventories on seven key building blocks underlying assessment methods (as outlined by Lievens & Sackett, 2017).

### **Research results**

Several studies have examined the links between self-report personality ratings and SJT scores, with results showing that SJTs can serve as alternative assessments of personality traits. For example, Mussel et al. developed an SJT for assessing five narrow traits (e.g. compliance, gregariousness, and self-discipline) and obtained an average convergent validity of 0.59 (from 0.41 to 0.70) with corresponding self-report ratings of these facets, whereas the average discriminant validity with noncorresponding facet ratings was  $-0.01$  (from  $-0.31$  to  $0.19$ ). The facet SJT scores also predicted relevant criteria over

<sup>3</sup>Apart from the 'What would you do?' instruction, a 'What should you do?' instruction ('What is the most effective action?') is often used. Yet, the relations between SJT scores with this instruction and personality ratings are lower (McDaniel, Hartman, Whetzel, & Grubb, 2007).

Table 1. Comparison of self-report personality inventories, Situational Judgement Tests (SJTs), and assessment centre exercises

Building block (Lievens & Sackett, 2017)	Self-report personality inventories	SJTs (low-fidelity simulations)	Assessment centre exercises (high-fidelity simulations)
Stimulus format	Generic	Verbal descriptions of situations	Actual simulated situations, role-players
Stimulus presentation consistency	High levels of standardization	High levels of standardization	At best, medium levels of standardization (predetermined cues are built in the exercises)
Contextualization	Low levels of contextualization	Medium levels of contextualization (brief, one-paragraph descriptions of task, characters, etc.)	High levels of contextualization (detailed descriptions of task, characters, etc.)
Content targeted	Behavioural tendencies ('typical performance')	Procedural knowledge about effectiveness of traits in situations	Actual trait-related behaviour in simulated situations ('maximal performance')
Response format	Self-reports	Multiple-choice responses	Open-ended responses
Response evaluation consistency	Trait scoring	Subject matter experts determine scoring key a priori (trait or effectiveness scoring).	Trained assessors make behavioural observations and ratings on dimensions.
Information source	Target person	Target person	Trained assessors

Note. The characteristics describe the prototypical version of the different methods, thereby acknowledging that variants exist.

and above the self-report ratings of the same facets. More generally, McDaniel et al. (2007) conducted a meta-analysis of SJTs with 'What would you do?' response instructions and found that SJTs had incremental value for predicting job performance over and above self-rated personality ( $R^2$  increments .06 and .07). So, while 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 behaviour. 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.

Motowidlo et al. (2006a, 2006b) proposed the notion of dispositional fit to explain why people's responses to SJT items show substantive correlations with analogous self-reported personality ratings (Motowidlo & Beier, 2010; Motowidlo et al., 2006a). In particular, they posited that people who possess high levels of a given trait are more likely to endorse an SJT response option that represents an action expressing that trait than people who have lower levels. In addition, Motowidlo et al. (2006a) argued that people who endorse SJT response options representing actions expressing specific traits will show corresponding trait-related behaviour more often in real life. In other words, the notion of dispositional fit posits that by endorsing response options in SJTs, people reveal something (i) about their standings on the personality traits represented by these response options and (ii) their likelihoods to show behaviour expressing these associated traits in real-life situations (Weekley, Hawkes, Guenole, & Ployhart, 2015).

To test these ideas, Motowidlo and colleagues measured implicit trait policies (ITPs; Motowidlo et al., 2006a) in SJTs requiring candidates to rate effectiveness of all response

options. ITPs connect two pieces of information: (i) the *target person's* effectiveness judgements of the response options and (ii) the trait levels of these response options as predetermined by *subject matter experts*. As shown in Appendix 1, a target person's ITP for a particular trait is then computed by correlating the target person's judgements of the effectiveness of response options with the predetermined trait levels of the response option. The correlation coefficient is used as an index of effect size in that it scales the degree to which variability in the target person's judgements about effectiveness of response options relates to variability in response options' trait expression (in agreeableness, for instance) across situations.

Given that ITPs are statistically computed instead of explicitly measured, ITPs scale how much the target person's judgements about the effectiveness of response options relate to the predetermined trait levels that these response options express, without telling the target persons that the response options they are judging vary on the trait in question (as noted in the preceding text, these trait levels are determined beforehand by subject matter experts). According to Motowidlo et al. (2006a), 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, they viewed them as characteristic adaptations instead of basic personality tendencies (see McCrae & Costa, 1996).

In several studies, Motowidlo et al. (2006a) found support for their ideas regarding the role of ITPs associated with agreeableness. For example, they examined whether ITP scores for agreeableness correlated with (i) self-reported agreeableness ratings on the NEO Personality Inventory and (2) behavioural 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 intended to activate agreeableness but were not exactly the same as the situations in the SJT. Independent observers rated degree to which participants expressed trait-related behaviours in the role-plays. They found that ITP scores for agreeableness significantly correlated with self-reported agreeableness *and* significantly predicted agreeableness scores in the role-plays ( $r$  between .23 and .45) over and above self-reported agreeableness.

### Differences between and similarities with existing personality approaches

Like social intelligence tests (e.g. Weis & Süß, 2005), emotional intelligence tests (e.g. MacCann & Roberts, 2008), and situation-response inventories (Furnham & Jaspars, 1983), SJTs present situation descriptions and ask people for the most appropriate responses. The scoring of these tests and SJTs is also similar, although the former are more often scored by comparing respondents' answers to what the general population would do (also known as consensus scoring), while SJT scores typically reflect the views of people with considerable knowledge and experience with the particular contexts and subject matter (expert scoring).<sup>4</sup> Moreover, the notion and computation of ITPs differentiate SJTs from these earlier measures.

At first sight, ITPs might resemble the classic personality concept of outcome expectancies (i.e., subjective associations between specified behavioural actions and outcomes) that plays a key role in various theories (e.g. social learning behaviour regulation and expectancy-value). However, the ITP notion is different because—as shown in Appendix 1—it statistically captures the target person's *policy* on the basis of that person's judgements of a *range* of behaviours (i.e., the policy is inferred on the basis of correlation between trait levels expressed by multiple ways of acting and effectiveness of those acts), not just expectancies regarding *individual* behaviours. So, ITPs do not refer to beliefs that something good or bad will happen if single specific acts are performed.

In conclusion, provision of scenarios to assess a target person's standing on personality traits in the form of SJTs represents an approach to go beyond traditional self-report personality inventories. Although SJTs resemble some existing measures and concepts in the personality domain, they also have unique aspects. These allow addressing important issues in personality research, which I discuss next.

### Opportunities for better integrating SJTs in personality psychology research

In the succeeding texts, I outline how SJTs might offer new avenues for assessing within-person variability and personality disorders. Moreover, I discuss how the procedural knowledge captured by ITPs opens important

opportunities for further illuminating trait–behaviour links. Table 2 summarizes these future research avenues of the use of SJTs in personality psychology. To put this agenda into practice, personality researchers are invited and encouraged to tailor the SJT approach to their own needs. There is no reason the SJT context should be work-related. Items with situations related to school, health, romantic relationships, sports, etc. could also be developed.

### Toward more control in testing explanations of within-person variability

Experience-sampling studies have shed important light on within-person variability (Wrzus & Mehl, 2015), typically evaluating both indices of within-person variability and mean scores across various occasions (e.g. Fleeson & Gallagher, 2009; Furr, 2009). However, experience-sampling research is not without drawbacks. Most notably, 'the lack of situational control is a potential weakness. The fact that participants likely encounter different situations, combined with the fact that behaviour is affected by situational stimuli, produces thorny complexities' (Furr, 2009, p. 385; see also Geukes et al., in press). Recently, Fleeson and Law (2015) addressed this concern by using a novel approach. They took experience-sampling studies to the lab: People were invited to come to the lab on repeated occasions to perform a series of tasks, while being observed by judges. This clever design presented the same situations/tasks to people in the same order, indicating to what degree consistency in behaviour was due to the actor or the situation.

I posit that SJTs are straightforward and cost-efficient methods that permit accomplishing similar objectives for two reasons. First, SJTs allow computing both means and variances in personality trait scores across the SJT situations (Dalal et al., 2015). Second, unlike experience-sampling field studies [but similar to Fleeson & Law's (2015) lab study], SJTs 'standardize' the situation component (i.e., everyone is presented with the same set of situations in the same order). Accordingly, SJT indices of within-person variability might help disentangle some of the usually confounded reasons for within-person variability (see also Geukes et al., in press). For instance, does high variability in extraversion reflect (a) exposure to or selection of varied situations calling for different levels of extraverted behaviours (a situation effect: i.e., anyone would show this much variability in extraversion in similarly varied situations), (b) sensitivity to contextual cues and ability to adapt behaviour accordingly (consistent with a behavioural signature of personality interpretation: Mischel & Shoda, 1995), or (c) capricious and unsystematic behaviour across situations (Baird, Le, & Lucas, 2006; Erickson, Newman, & Pincus, 2009)? As SJTs standardize the situation, explanation (a) is eliminated, although ecological validity is reduced due to elimination of individual choice in entering situations.

If one is interested in correlating SJT scores with performance outcomes (as often done in selection) or with measures of life adjustment (as often done in personality research), it is possible to disentangle the other explanations too. That is, positive or higher correlations between intraindividual variability and desired outcomes suggest that

<sup>4</sup>All of these tests also rely sometimes on a hybrid of consensus and expert scoring.

Table 2. Summary of future research directions for better integrating Situational Judgement Tests (SJTs) in personality research

General recommendations for research on SJTs and personality
<ul style="list-style-type: none"> <li>• Develop and validate SJT items that target personality factors and facets (e.g. Big Five and HEXACO)</li> <li>• Develop SJT items for domains other than work (e.g. school, health, relationships, sports, and personality disorders)</li> <li>• Validate SJTs for these different domains and factors/facets so that they consist of sufficient items and establish convergent and discriminant validity with existing self-report personality measures</li> <li>• Develop SJTs with higher ecological validity such as SJTs with open-ended responses (e.g. oral, written constructed, and webcam)</li> </ul>
Advancing research on within-person variability via SJTs
<ul style="list-style-type: none"> <li>• Determine the number of SJT items needed to obtain reliable within-person variability SJT scores</li> <li>• Use SJTs as a different method to test the key role of within-person variability as suggested by whole trait theory</li> <li>• Examine convergence of within-person variability as obtained via an SJT approach with within-person variability as obtained via an experience-sampling approach</li> <li>• Use SJTs to disentangle key reasons (situation effect, person–situation effect, etc.) behind within-person variability</li> <li>• Develop SJT items on the basis existing situational taxonomies (e.g. DIAMONDS and CAPTION) and use such SJTs to uncover situation–trait contingencies as suggested by the cognitive affective personality system (CAPS)</li> </ul>
Advancing research on personality disorders via SJTs
<ul style="list-style-type: none"> <li>• Examine convergence of SJTs that assess personality disorders with existing personality disorder measures</li> <li>• Examine incremental validity of SJTs that assess personality disorders over and above existing personality disorder measures for predicting criterion data</li> <li>• Examine convergence of within-person variability as obtained with SJTs that assess personality disorders to within-person variability as obtained with an experience-sampling approach</li> <li>• Use people’s responses and construal to close-ended and/or open-ended SJT items that assess personality disorders to test the role of encoding biases in theories of personality disorders</li> <li>• Use people’s responses and construal to close-ended and/or open-ended SJT items that assess personality disorders as basis for individual treatment sessions and cognitive behavioural therapy exercises</li> </ul>
Advancing research on ITPs and mediators between traits and behaviour via SJTs
<ul style="list-style-type: none"> <li>• Examine convergence of people’s implicit trait policies (ITPs) about specific traits with external beliefs about trait effectiveness in situations and with other implicit measures</li> <li>• Examine convergence of people’s ITPs about specific traits with self-reported and peer-reported traits</li> <li>• Examine whether people’s ITPs about specific traits predict their simulated and actual behaviour over and above self-reported and peer-reported traits</li> <li>• Examine the role of ITPs and other mediation factors in the link between traits and trait-related behaviour in actual situations.</li> </ul>

variability is adaptive [see explanation (b)], whereas negative or lower correlations with desired outcomes indicate the opposite [explanation (c)]. Lievens et al. (in press) set up such a design in the selection domain. They developed a 24-item SJT and 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 ( $R^2$  change = .05) as well as in actual personality state variability ( $R^2$  increments from = .04 to .07) measured 2 years later via a 10-day experience-sampling study.

Apart from standardizing situations, SJTs offer opportunity to examine situation–trait contingencies. A challenge, however, is that recent research shows that situations have less impact on SJT responses than previously thought (Krumm et al., 2015). As the scenarios of most SJTs are typically not developed on the basis of any theoretical framework of situations, it would be fruitful for selection and personality researchers to collaborate in deliberately manipulating situational characteristics built into SJT items on the basis of recent situational taxonomies (e.g. Parrigon, Sang, Tay, & Wang, 2017; Rauthmann et al., 2014). For example, using each of the eight situational dimensions in Rauthmann et al.’s DIAMONDS framework, a set of SJT-style item stems could

be designed to examine to what extent people’s (between-person) or individuals’ (within-person) responses differ across and within situations that tap into specific situational dimensions. This would permit examining how people’s response option choices systematically vary according to specific situational dimensions or cues so that meaningful group and individual variation in situation–trait contingencies can be studied (Huang, Ryan, Zabel, & Palmer, 2014). Such ‘if–then’ patterns are inherent parts of theoretical frameworks such as cognitive affective personality system (Mischel & Shoda, 1995), and SJTs that are developed on the basis of situational taxonomies enable assessing these situation–trait contingencies and whether these are (mal)adaptive.

Despite these opportunities, some methodological caveats are in order. Within-person variability can be examined only when SJTs consist of sufficient numbers of items (Wang & Grimm, 2012) and indeed assess the personality traits targeted, but both seem manageable tasks. For example, Mussel et al. (2016) obtained good convergent and discriminant validities for SJT facet scores on the basis of 22 SJT items per personality facet. In addition, SJTs’ greater situational control in measuring within-person variability comes with a price. SJTs are ecologically less valid than experience-sampling studies, and their scores do not reflect that people select much about how and whether

to enter the situations they experience, nor that personality is involved in this selection (Ickes, Snyder, & Garcia, 1997; Rauthmann, Sherman, Nave, & Funder, 2015). As noted in the preceding texts, SJTs are based on the assumptions that people encounter similar situations and that their knowledge about appropriate actions in these situations predicts future trait-related behaviour more generally (Lievens & Patterson, 2011; Lievens & Sackett, 2012; Motowidlo et al., 2006a).

Future research is therefore needed to take up these methodological challenges. As one potential solution, SJTs have been developed that replace the multiple-choice option format with open-ended responses (e.g. Rockstuhl, Ang, Ng, Lievens, & Van Dyne, 2015). In addition, branched or nonlinear SJTs (e.g. Olson-Buchanan et al., 1998), in which candidates' responses to earlier items determine the subsequent situations that they receive, reflect the notion that people influence which situations they enter. Moreover, it is possible to increase ecological validity by requiring respondents to enact their responses to a video-based situation in front of a webcam (e.g. Cucina, Chihwei, Busciglio, Harris Thomas, & Thompson Peyton, 2015; Lievens, De Corte, & Westerveld, 2015a). These webcam vignettes are then coded and rated by experienced assessors.

### **Toward additional instruments and insight related to personality disorders**

Recent research has investigated short-term fluctuations in manifestations of personality disorders in daily life (see Ebner-Priemer, Eid, Kleindienst, Stabenow, & Trull, 2009; Santangelo, Bohus, & Ebner-Priemer, 2014; Trull et al., 2008; Wright & Simms, 2016). For example, Wright and Simms (2016) followed a group of individuals diagnosed with any personality disorder (e.g. paranoid, schizoid, schizotypal, and borderline) for 100 days and found substantial daily variability but stable monthly averages and levels of variability. Their study used experience-sampling in real-life settings. Although ecologically valid, it is more difficult to use this method in clinical diagnosis, for which inventories and interviews still predominate.

Situational Judgement Tests might provide an alternative method in the context of the assessment of personality disorders. Compared with existing inventories, SJTs have at least two advantages. For one, they specify situations. In addition, they enable computing mean scores across situations to quantify between-person variance as well as situation-specific individual deviation scores to quantify within-person variance (provided that they have a sufficient number of items and that relations between response options and symptom/trait levels have been established beforehand).

To examine the viability of SJTs for assessing personality disorders, an SJT could present a wide variety of life situations on the basis of existing situational taxonomies (such as DIAMONDS or CAPTION; see Rauthmann et al., 2014; Parrigon et al., 2017). The response options to these situations could reflect various pathological and typical personality indicators (e.g. Krueger, Derringer, Markon, Watson, & Skodol, 2012; Simms et al., 2011). Asking a group of individuals diagnosed with any personality disorder and a

community sample to complete such an SJT and traditional personality disorder inventories would allow examining how SJT scores perform relative to already established inventories. By including criterion data (e.g. measures of distress or social or occupational dysfunction), one could examine whether SJT mean scores account for incremental variance over and above established inventories in these criteria. Furthermore, correlations between intraindividual variability scores SJT and actual variability in personality states gathered via an experience-sampling study (see Wright & Simms, 2016) could be investigated to estimate whether such SJTs are viable methods for uncovering fluctuations in personality disorder manifestations.

Using SJTs as alternative measures of personality disorders has not only methodological appeal. SJTs that assess personality disorders might also be used to test the roles of encoding biases and cognitive distortions in personality disorders. One might examine how a person psychologically construes various situational features in SJT items differently (e.g. more threateningly) than how they are consensually perceived. Although closed (multiple-choice) formats might be used, reliance on open-ended formats in which people formulate their own response might be fruitful, especially in initial efforts. The SJT approach therefore has also potential for informing treatment approaches. One might inspect to which situations a particular person responds most 'strangely' and explore how the person construes them. On the basis of this information, cognitive behavioural therapy exercises might be developed to train the person to respond in other ways or to avoid such situations.

### **Toward testing ITPs as potential mediators between traits and behaviour**

Implicit trait policies have emerged as one of the innovative notions in recent SJT research. So far, empirical evidence has been established for associations between people's trait levels (determined on the basis of self-reports on personality inventories) and ITP scores (obtained via SJTs) and between ITP scores and behaviour in simulated situations (role-plays) activating the trait of interest. In addition, ITP scores explained incremental variance in trait-related behaviour over and above self-reported trait levels.

Personality researchers can put the ITP construct to further test. First, it is important to gather more evidence regarding the roles of ITPs related to a variety of traits. So far, evidence has primarily been gathered for ITPs associated with agreeableness and related traits (e.g. prosociality). Studies are also needed to establish that people actually possess implicit versions of the posited policies, as well as their convergence with other explicit and implicit beliefs. Explicit beliefs, posited by social-cognitive theories, might be captured by self-reports. Implicit beliefs are typically more difficult to measure but might, for instance, be obtained via implicit association tasks. Other studies might go beyond self-reports as comparison standard and examine correlations between ITP scores and peer-reported traits (although the latter do not reflect intentions in behavioural expressions). Equally important, one should investigate to what extent ITP scores also predict actual



behaviour (apart from simulated behaviour in role-plays) and whether they do so over and above self-reported and peer-reported traits. Finally, one might ascertain whether ITP scores need to be computed on the basis of context-specific SJTs or whether, say, ITP scores derived from an SJT with sport club scenarios also predict friendship behaviour at school.

Second, consideration of ITP scores can enhance contemporary personality research on links between personality traits and trait-related behaviours in various situations. ITPs have considerable promise as under-researched potential mediators of links between traits and trait-related behaviour because ITP scores statistically scale people's judgements of the effectiveness of trait-related behaviours in situations that are hypothesized to act as precursors of actual behaviour in such situations. More generally, a comprehensive investigation could be set up in which other potentially mediators of the trait-behaviour link are examined alongside ITP scores. For example, people might also be more likely to display traits because of (i) prior investments in acquiring skill in displaying particular trait-related behaviours, (ii) self-concepts (ideas of what kind of person they are and thus how they want to act), or (iii) greater salience of competing behavioural cues to feature other traits.

As an example study, one might examine the role of ITP scores in various domains (e.g. school, work, and sport club) for predicting prosocial behaviour or in interpersonal perception (e.g. reputation). This would require developing specific SJTs for each of these domains and computing ITP scores for the traits considered relevant. This basic design might be extended by adding measures of other factors contributing to behavioural displays and examining their relations. Experience-sampling data and other reports of people's behaviour could serve as criterion variables. Results from the various domains could be scrutinized to identify to what degree the various factors (ITPs, prior investments in acquiring skill in displaying particular trait-related behaviours, etc.) mediate trait-behaviour links.

## ASSESSMENT CENTRE EXERCISES

### Assessment centre exercises: Definition, characteristics, and scoring

In personnel selection, there is a long tradition of providing insight into candidates' personalities via behavioural observation in actual stimulated situations (Thornton & Cleveland, 1990). Assessment centre methodology typifies this rich history. In assessment centres, candidates participate in multiple simulation exercises (e.g. role-plays, presentations, and group discussions), while independent trained observers observe and rate their behaviour. Thus, candidates have the opportunity to demonstrate that they can behave appropriately in a simulated workplace situation. That is one of the reasons that job candidates tend to have favourable perceptions of these exercises (Hausknecht et al., 2004).

The first assessment centres (before WWII) were developed in military contexts (officer selection in Germany, UK, and USA), lasted at least 2 days, and consisted

mainly of group exercises. They were considered effective ways to obtain 'holistic views' of candidates' personalities (Gibbons & Rupp, 2009). The assessment centres that became popular in the 1970s and 1980s for managerial selection operated similarly, although their exercises were shorter in number and duration, and introduced mixes of individual, one-on-one, and group activities. Given that these exercises are expensive to design, there has also been a strong trend toward using them not only for hiring/selection purposes but also for developmental purposes (e.g. to provide detailed feedback about participant strengths, weaknesses, and suggestions for required training). In such 'developmental' assessment centres, it is not uncommon that the independent observer (assessor) ratings are supplemented by self-ratings and peer-ratings from fellow assesseees to obtain more comprehensive pictures of participants. Recently, remote/online versions have also been developed to reduce expense and increase convenience (assessment at any time and in any place). Another modern alternative is the multiple 'speed' assessment centre in which candidates participate online or in real life in a large set of (e.g. 18) short (e.g. 3-minute) role-plays (Herde & Lievens, 2016).

Assessment centre exercises are considered high-fidelity simulations because they provide detailed organizational, social, and task information (Tett & Burnett, 2003) and require people to respond with actual behaviour. In the exercises, the participants receive pages of information about the characters of the other people in the situation, the roles they play in it, the firm, the problems that need to be tackled, etc. Such information is intended to enable candidates to fill their roles in the simulated situations realistically (Thornton & Rupp, 2006). Situation outcomes emerge during the role-playing, so candidates must generate their own approaches instead of choosing among specified response options as in SJTs. In some exercises (e.g. group discussions), the other participants are fellow candidates, whereas in other exercises (e.g. a presentation to a board), the other role-players have assignments about how to act.

Assessors are taught to adhere to a data-driven information processing model in which they distinguish observation from evaluation: They receive specific training (from 1 day to up to a week) in which they learn to begin by observing and noting in as much detail as possible the (non)verbal behaviours of the participants. They are taught to classify their observations next into prespecified dimensions (competencies) that are deemed to be important to the job at hand. These dimensions (e.g. communication and sensitivity) reflect bundles of behaviours that the employers require job holders to display readily and skillfully. Arthur, Day, McNelly, and Edens (2003) collapsed 168 assessment centre dimensions into a taxonomy of seven meta-dimensions: influencing others, communication, drive, tolerance for stress, consideration/awareness of others, problem-solving, and organizing and planning. Only after observing, noting down, and classifying behaviours are assessors expected to evaluate the candidates on these dimensions. As the situations are representative of typical job situations, in the evaluation stage, the assessors are evaluating whether the candidates performed well on the dimensions considered relevant and thus whether



the kinds and levels of behaviour displayed fit the job and culture of the organization. Research reveals that assessor training is generally effective in aligning assessor rating processes and ensuring adequate inter-rater reliability (Thornton & Rupp, 2006). Table 1 summarizes the characteristics and scoring of assessment centre exercises and compares them with personality self-reports and SJTs.

### Research results

Similar to SJTs, there is a voluminous research base on relations between assessment centre ratings and self-report personality ratings, including two meta-analyses (Hoffman, Kennedy, LoPilato, Monahan, & Lance, 2015; Meriac, Hoffman, Woehr, & Fleisher, 2008). Meriac et al. (2008) examined relations between candidates' personality self-reports and typical assessment centre dimension ratings provided by assessors, whereas Hoffman et al. (2015) investigated links between candidates' self-reported personality ratings and assessors' overall effectiveness ratings in the exercises. In both meta-analyses, correlations were at best moderate (in the .20s to .30s), which is lower than typical correlations between self-ratings and other ratings of personality (Connelly & Ones, 2010). Importantly, however, similar to SJT research, this meta-analytic research indicated that assessment centre ratings had incremental validity in predicting job performance over and above self-report general personality ratings and general cognitive ability scores ( $R^2$  change between 3% and 10%), likely due to their greater relevance to specific context.

The at-best moderate correlations between assessment centre ratings and self-report personality are not surprising for several reasons. First, assessors are not rating personality traits as such but candidate effectiveness in situations, even though some rated dimensions are conceptually related to personality traits (e.g. stress tolerance to emotional stability). Personality research linking personality self-reports to actual behavioural ratings on single dimensions found similar correlations (e.g. Back, Schmukle, & Egloff, 2009; Borkeau, Mauer, Riemann, Spinath, & Angleitner, 2004; Funder, Furr, & Colvin, 2000; Gosling, John, Craik, & Robins, 1998; Leikas, Lönnqvist, & Verkasalo, 2012; Vazire & Mehl, 2008). Second, in assessment centre exercises, performance tends to be optimized (also known as 'maximal' compared with 'typical' performance) because of the high-stake consequences for candidates. This might especially impact on the validities for motivationally relevant personality traits such as conscientiousness and achievement motivation. Third, the range of personality traits in the applicant pool is typically restricted to the extent that applicants gravitate toward the job in question based on personality traits, which might produce lower correlations than would be observed in the population at large. Fourth, as assessment centre exercises essentially rely on observation of behaviour, traits that manifest less easily in behaviour or have less salient behavioural cues might be more poorly assessed or evaluated than others (see Vazire, 2010). The reverse is possible for more easily observable dimensions. For example, there is evidence that oral communication can be

rated very well in assessment centre exercises (Bowler & Woehr, 2006).

### Differences between and similarities with existing personality approaches

Assessment centre exercises depart in two major ways from classic self-report measures: (i) detailed, rich, and interactive situations and (ii) focus on actual behaviour (instead of self-reported behavioural tendencies), although in simulated settings. Assessment centre designs do share resemblances to the literature on accuracy of zero/short-acquaintance personality assessments (e.g. Borkeau et al., 2004; Hirschmüller, Egloff, Schmukle, Nestler, & Back, 2015) and to various theoretical models [e.g. Funder's (1999) realistic accuracy model (RAM) and Kenny's (2004) PERSON model]. So far, however, the assessment centre and zero/short-acquaintance personality assessment literatures have evolved mostly independently from each other.

Similarities between assessment centre and the zero/short acquaintance are that, in both these domains, assessors are independent observers who are typically not acquainted with the target persons and judge them in mostly interpersonal contexts. In fact, the RAM stages can be framed in assessors' observation and coding processes in assessment centre exercises. For example, assessors' notes can be scrutinized to determine the detection of relevant and available behaviours, whereas their dimension assignments can serve as indicators of utilization. Furthermore, assessment centre research speaks to the moderators (good targets, good judges, good information, and good traits) of the RAM.

Differences between the two literatures' approaches are that assessment centre candidates are highly motivated to perform well (compared with typical research participants in lab studies on zero/short acquaintance), assessors have specific training and are looking for specific behaviours (instead of research participants rating each other in round-robin designs according to often idiosyncratic criteria), assessment centres work with specific job-related dimensions (instead of personality traits), and the exercises last longer than many brief 'get-to-know-you' tasks in zero/short acquaintance research. The dependent variables are also different. Whereas in zero/short acquaintance studies self-other agreement indices are typically used as operationalizations of accuracy, the predictive validity of the ratings is often viewed as the most important outcome in the selection literature. That said, to determine accuracy, both literatures also rely on videotaped targets with known trait levels and objectively coded behavioural cues.

I view these differences between the two literatures more as opportunities than as barriers for more integration between them. Given their somewhat different focus, these two literatures complement each other well. In the next section, I therefore discuss how these two strands of research might be better integrated and cross-fertilize each other so that in the end, richer understandings of personality expression and perception are obtained.

### Opportunities for better integrating assessment centre exercises in personality psychology research

This section has three main parts. I start with arguing that personality research might draw on assessment centre research about the impact of situation–trait relevance and situational strength on personality expression. Additionally, I posit how personality researchers might fruitfully build on recent developments on personality perception (dispositional reasoning) in the assessment centre and interview domains. Finally, I argue that recent focus on transactions among the different RAM stages in assessment centre research offers opportunities for future research on the interplay between personality expression and interpersonal perception. Table 3 summarizes the various future research avenues in personality psychology that might be tackled via assessment centre exercises.

Similar to SJTs, I stress that personality researchers can tailor the assessment centre approach to their own needs. For example, one can specify other rating dimensions<sup>5</sup> and other situations besides work environments. In fact, in lab settings of personality studies, actual assessment centre exercises or tasks that have close resemblance to assessment centre exercises (e.g. Back et al., 2009; Fleeson & Law, 2015; Gosling et al., 1998; Kirschbaum, Pirke, & Hellhammer, 1993; Robins & Beer, 2001) have already been used, thereby underscoring their applicability in personality research.

### Toward better insight in personality expression

Trait activation theory (Lievens, Tett, & Schleicher, 2009; Tett & Burnett, 2003; Tett & Guterman, 2000) has emerged as an important framework in the assessment centre field for better understanding trait expression. Building on early works by Murray (1938) and Allport (1951), trait activation theory addresses how individual traits come to be expressed as behaviour in response to trait-relevant situational demands. Two factors are posited to be of central importance. First, *situation–trait relevance* refers to the extent to which qualitative situation features provide particular cues that increase likelihoods of people displaying more behaviours relevant to a given trait than to another trait. Such cues are considered to fall into three broad and interrelated categories: task/individual, social/group, and the broader context. Tett and Burnett (2003) offered a wide array of examples of cues at these different levels that might activate trait-related behaviours (e.g. a messy cabinet evoking orderliness-related behaviour, see their Table 1, pp. 508).

Second, *situation strength* refers to clarity and imperative nature of situational cues. A strong situation produces similar behavioural responses from virtually all individuals, whereas responses vary considerably in weak situations (Meyer, Dalal,

& Hermida, 2010; Mischel, 1973). For example, cues triggering interpersonal sensitivity might vary from someone's momentarily distressed facial expression to overt sobbing. Situational strength is considered high when demands are consistent and clear and have important positive or negative social consequences and when appropriate responses fall within narrow ranges (see framework of Meyer et al., 2010).

In the assessment centre domain, trait activation theory has been used to identify which assessment centre exercise features trigger candidate behaviour. To this end, pretested and predetermined cues have been planted at the task, social, and context levels in the exercises' content/instructions and in role-player assignments (see Lievens, Schollaert, & Keen, 2015b; Oliver, Hausdorf, Lievens, & Conlon, 2016; Schollaert & Lievens, 2012). For example, stringent time limits, sudden obstacles, and information overload have been used to evoke how candidates deal with stress.<sup>6</sup>

Although trait activation theory proposes conditions that affect expression of trait-related behaviour, results so far show at best moderate support and mostly trait–behaviour intent relations (instead of trait–behaviour relations) have been studied by using written scenarios. Tett and Guterman (2000), for instance, found that people tended to express greater intent to express particular traits for posed situations cuing those traits (highest average correlation was .46 across situations) than for situations that did not cue those traits. Yet, these trait-behavioural intent relations were significant in only one-third of the scenarios. A large-scale study that focused on candidate performance (instead of on behavioural intentions) further revealed that consistency of candidate performance on the same dimensions across exercises was only slightly higher when assessment centre exercises were similar in situation–trait relevance (Lievens et al., 2006). A recent experience-sampling study (Sherman, Rauthmann, Brown, Serfass, & Jones, 2015) reported similarly small effect sizes for statistical trait–situation interactions as predicted by trait activation theory.

While these results offer at least some support for trait activation theory, they also highlight that any situation can elicit expression of many different traits to varying degrees, and people can behave similarly though motivated by different traits. Or as Tett and Guterman (2000, pp. 413–414) put it: 'Real-world situations contain cues relevant to multiple traits. Even in an apparently simple case where one trait seems to dominate in relevance, cues relevant to other traits can interfere, making responses in that situation difficult to predict.'

Thus, although conceptual and empirical progress related to trait activation has been made in the assessment centre field, more work is definitely needed. This is an area in which personality and selection researchers could work together. First, there is need to scrutinize measurement of situation–trait relevance and situation strength. So far, subject-matter experts have typically made situation strength and trait relevance ratings. For instance, this procedure was used not only for assessment centre exercises (e.g. Lievens et al., 2006) but also in the context of trait–job performance relationships across different jobs (Judge & Zapata, 2015). The means of the subject matter expert ratings are then

<sup>5</sup>There is also research that links the seven basic assessment centre dimensions to personality traits via subject matter experts (Lievens, Chasteen, Day, & Christiansen, 2006) and some assessment centres even incorporate trait judgements (see Christiansen, Hoffman, Lievens, & Speer, 2013; Lievens, De Fruyt, & Van Dam, 2001).

<sup>6</sup>Use of such predetermined cues to elicit trait-related behaviour bears resemblance to the Trier Social Stress Test (Kirschbaum et al., 1993) in which actual or virtual role-players are scripted to evoke behaviour.

Table 3. Summary of future research directions for better integrating assessment centre exercises in personality research

General recommendations for research on assessment centre exercises and personality
<ul style="list-style-type: none"> <li>• Develop assessment centre exercises for domains other than work (e.g. school, health, relationships, and sports)</li> <li>• Examine how typical assessment centre dimensions relate to personality traits</li> <li>• Validate assessment centre exercises to establish convergent and discriminant validity with existing self-report and peer-reported personality measures</li> <li>• Develop assessment centre exercises on the basis existing situational taxonomies (e.g. DIAMONDS and CAPTION) and use such exercises to examine their trait activation potential.</li> </ul>
Advancing insight in personality expression via assessment centre exercises
<ul style="list-style-type: none"> <li>• Use an assessment centre exercises (e.g. cues provided by role-players) to further identify and validate cues at the task, group, and broader context level that trigger trait-related behaviour</li> <li>• Examine how consensual measures of situation–trait relevance compare with psychological ones in predicting trait-behaviour links and behavioural consistency across situations</li> <li>• Examine the combined influences of multiple (competing) cues on trait expressions</li> <li>• Use assessment centre exercises (instructions given to candidates) to examine the effects of situational strength on the behavioural options between persons, the behavioural option within persons, the between-person variance in the level of one given behaviour, and the within-person variance in the level of one given behaviour across similar instances of the same situation</li> <li>• Examine what motivates people to behave in a particular way in the assessment centre exercises. Is this cognition-related (e.g. ITPs and other implicit and explicit beliefs), trait-related (as suggested by trait psychological theories), goal-related (as suggested by approach/avoidance goals), or impression-related (impression management theories)?</li> </ul>
Advancing insight in personality perception via assessment centre exercises
<ul style="list-style-type: none"> <li>• Further test the psychometric properties (reliability, convergent, and discriminant validity) of the dispositional reasoning measure in a variety of contexts and for a variety of traits</li> <li>• Validate the dispositional reasoning measure with policy capturing procedures typically applied in Brunswikian approaches</li> <li>• Examine the added predictive value of dispositional reasoning over and above the individual differences (e.g. personality traits and acquaintance level) commonly examined to understand interpersonal perception in a variety of contexts</li> <li>• Examine dispositional reasoning as one of the reasons behind low self-other agreement</li> </ul>
Advancing insight in the interaction between personality expression and personality perception via assessment centre exercises
<ul style="list-style-type: none"> <li>• Specify circumstances and concrete tasks that increase/decrease accuracy differences among judges</li> <li>• Examine the interaction between the ‘good judge’ and the ‘good trait’ factors of the RAM on accuracy and accuracy differences among judges</li> <li>• Examine the interaction between the ‘good judge,’ the ‘good information,’ and ‘good target’ factors of the RAM on accuracy and accuracy differences among judges</li> </ul>

typically used as objective measures of situation–trait relevance [also known as canonico-consensual measures (Block & Block, 1981) and alpha press (Murray, 1938)]. It is equally important to gather participants’ individual perceptions of situation–trait relevance [also known as subjective–psychological measures (Block & Block, 1981) and beta press (Murray, 1938)] and examine how consensual measures of situation–trait relevance compare with psychological ones in predicting trait–behaviour links and behavioural consistency across situations (see Sherman, Nave, & Funder, 2010). Similarly, assessment centre research has examined the trait-activation potential of situations (exercises) of only a limited number of cues and traits. So, the combined influences of multiple (competing) cues on trait expressions are in dire need of research.

Second, the strength of the instructions (e.g. detailed versus vague instructions) given to candidates in assessment centre exercises can be used to conduct an in-depth examination of the effects of situational strength. A strong situation can restrict or constrain several aspects<sup>7</sup>: (i) the behavioural options

between persons (e.g. only certain behaviours are permissible), (ii) the behavioural options within persons (e.g. only certain behaviours from a person’s behavioural repertoire will be selected); (iii) the between-person variance in the level of a given behaviour (e.g. most people behave similarly highly on one behaviour in that situation), and (iv) the within-person variance in the level of one given behaviour across similar instances of the same situation (e.g. the person enacts the same level of the same behaviour in every instance of the situation, above and beyond his or trait level).

Finally, selection and personality researchers could work together to examine how typical assessment centre dimensions relate to personality traits and how particular ‘solutions’ or behavioural displays reflect particular personality traits. Regarding the latter, a comparison between assessment centre exercises for high-stake (selection and promotion) versus low-stake purposes (self-development) allows examining the role of various drivers of what motivates people to behave in a particular way in the exercises (see Sherman et al., 2015). The drivers examined might be cognition-related (e.g. ITPs and other implicit and explicit beliefs), trait-related (as suggested by trait psychological theories), goal-related (as suggested by approach/avoidance goals), or impression-related

<sup>7</sup>I thank John Rauthmann for these insights.



(impression management theories). Impression-related aspects might dominate in the high-stake setting, whereas they might play less important roles in the low-stake developmental setting. Such a study is informative to which extent the drivers differ when the stakes involved differ (which is also relevant to research on the social effects of personality).

### Toward better insight in personality perception

Recent research provides new insights into what makes some assessors better judges (Christiansen, Wolcott-Burnam, Janovics, Burns, & Quirk, 2005; Lievens et al., 2009). These insights about the roles of individual differences in personality perception among observers result from interest in improving assessor and interviewer screening. In particular, Christiansen et al. (2005) introduced the idea of dispositional reasoning as declarative knowledge structures that enable processing of behavioural information. They argued that these knowledge structures involve three interrelated processes. First, trait induction refers to knowledge of how traits manifest themselves in behaviour. Trait induction helps a judge link behaviour cues to specific individual traits. Second, trait extrapolation denotes understanding of how traits and their behavioural manifestations naturally covary. This process permits inferring other traits likely involved in the target's personality profile. Third, trait contextualization is the ability to identify situations that are relevant in activating trait-related behaviours. Trait contextualization thus stresses the role of context in interpreting cues (Trope, 1986).

A key methodological advantage is that none of these components is measured through a self-report questionnaire. Instead, they are measured via multiple-choice tests in which people, for instance, have to assign adjectives to traits or determine which situation is best for observing specific trait-related behaviour (see Tett & Guterman, 2000, for examples). In addition, a measure of dispositional reasoning can be constructed with multiple-choice questions about any trait (so it is not restricted to the five-factor model), as long as the three processes are measured.

De Kock, Lievens, and Born (2015) observed that dispositional reasoning was the strongest predictor of accuracy (measured via a combination of self-reports and other reports) among individual differences including demographics and general measures of personality and cognitive ability and predicted incremental variance over those 'usual suspects.' This result is not without importance because cognitive ability typically shows the strongest relations with being a good judge in a variety of domains (e.g. Murphy & Hall, 2011).

Given that 'it has been so difficult to identify consistent predictors of being a good judge' (Human & Biesanz, 2013, p. 250), dispositional reasoning deserves a place on the agenda of personality psychology researchers.<sup>8</sup> To start with, selection and personality researchers could collaborate to improve the measurement of dispositional reasoning. The current measures provide either an overall score across the three components (Christiansen et al., 2005) and/or separate scores per component (De Kock et al., 2015). In light of potential

judge–trait interactions, constructing measures for assessing dispositional reasoning related to specific traits also deserves attention. Second, I suggest that personality researchers validate measurement of the dispositional reasoning components with policy-capturing procedures typically applied in Brunswikian (Brunswik, 1952; see also Nestler, Egloff, Kűfner, & Back, 2012) approaches. The overall aim of the dispositional reasoning measure is to operationalize cue utilization. To this end, the various subtests present cues (individual cues and cues built into short situations) to judges who have to assign a cue—relying on their internal schemas about how traits, behaviours, and situations go together—to a trait (or situation) category (depending on the particular subtest). Therefore, one might validate the dispositional reasoning measure by asking perceivers to rate videotaped targets with known trait values and objectively coded behavioural cues. Next, a cue sensitivity score that reflects per perceiver the match between actual cue validities and his/her cue utilizations could serve as a potential criterion. Third, dispositional reasoning deserves to be better integrated in zero/short acquaintance studies. For example, research could examine the added value of dispositional reasoning over and above individual differences commonly examined to understand interpersonal perception (e.g. in personality traits and acquaintance level; see Back & Nestler, 2016). Dispositional reasoning might also shed additional light on self-other rating discrepancies. That is, low self-other agreement might also result from people having weakly developed (or even incorrect) schemas related to traits, their indicators, and situations activating them.

### Toward better insight in the interplay between personality perception and expression

The utility of any approach focused on the good perceiver hinges on the question whether there is reliable accuracy variance between perceivers. Recent personality research (e.g. Human & Biesanz, 2013) has shown that there is not much variance between perceivers (much less than between targets), at least in differences in the utilization of available cues (see also Allik, de Vries, & Realo, 2016; Haselton & Funder, 2006). In selection, however, the rating task tends to be more challenging because candidates try to manage the impressions they create, making it more difficult for assessors to discern 'valid' from 'invalid' behavioural cues. In addition, assessors do not rate candidates solely on dimensions but also use these ratings to predict whether someone would perform well on the job and/or fit the organization. Hence, in selection, differences in predictive validity depending on the rater have been observed (e.g. Pulakos, Schmitt, Whitney, & Smith, 1996; Van Iddekinge, Sager, Burnfield, & Heffner, 2006) and rater training research has revealed that judges vary in accuracy prior to the training (e.g. Roch, Woehr, Mishra, & Kieszczyńska, 2012; Woehr & Huffcutt, 1994).

Given these diverging results in the zero/short acquaintance and selection literatures, an interesting contribution from assessment centre research could be to specify the circumstances and concrete tasks that

<sup>8</sup>I am indebted to Francois De Kock for suggesting some of these ideas.

increase/decrease accuracy differences among judges. Along these lines, recent assessment centre research has paid attention to relations between the good judge and other RAM moderators, although such relations have not received much attention in personality research. Conceptually, such studies contribute to integration of personality expression and perception models. As one example, future research could focus on transactions between the judge and the trait. One might examine whether results of ceiling effects of good judges' accuracy in zero/short acquaintance situations are also found when judges rate not only the personality of the target but also use these ratings to predict whether the target person fits in the peer group or is a good roommate (such behavioural prediction then mirrors predictive validity designs in selection).

As another example, in recent assessment centre, research efforts have been undertaken to elicit 'good information' among targets (i.e., judge-information and judge-target transactions) to make the rating task easier for assessors. For instance, Lievens, Schollaert et al. (2015b) integrated trait activation theory into the RAM. They conducted lab and field studies in which they manipulated two factors. One factor related to personality expression: situation-confederate behaviours intended as cues to activate trait-related candidate behaviours. The other factor dealt with interpersonal perception, specifically assessor familiarity with the specific cues. Most zero/short acquaintance personality research has considered only the first factor, finding it positively associated with accuracy. Interestingly, in contrast, Lievens et al. observed that cue availability affected only detection but neither utilization nor accuracy. Yet, when cue familiarity was considered and assessors were thus also familiar with them, utilization and accuracy significantly improved. This underscores the *interplay* between personality expression (behavioural elicitation by the confederates) and interpersonal perception (behavioural evaluation by the assessor). It also suggests more positive potential for training than most zero/short acquaintance studies do. In this study, training alerted assessors to the situation ('if') and the subsequent candidate reaction ('then'), thereby helping them to recognize potential situation-behaviour relations better. These results match well with scarce zero/short acquaintance research (e.g. Letzring, 2008) that finds more variance when perceivers can interact with targets, thereby pointing to the fact that accuracy differences might stem more from differences between perceivers in how much they are able to evoke behavioural cue differences in targets instead of from merely using the cue differences that are observable.

In short, to advance future research on personality expression and interpersonal perception, I suggest zero/short acquaintance researchers pay more attention to training approaches fleshed out in personnel selection and especially to the interplay among good judges, good traits, good targets, and good information. More generally, focus on transactions between judges and various RAM factors makes assessment centre research complementary to zero/short acquaintance research because such transactions have seldom been investigated in that domain.

## CONCLUSION

Undoubtedly, the personnel selection field has benefitted greatly from drawing on developments in the personality domain (Christiansen & Tett, 2013; Sackett et al., 2017). This article posited that two personnel selection approaches (SJTs and assessment centre exercises) that evolved relatively independently from the personality psychology field might also be fruitful for personality research. I therefore reviewed and discussed the potential of these two selection methods to be better integrated into future personality psychology studies. Incorporating these assessment approaches further broadens the already impressive methodological spectrum of techniques in personality research (see Wrzus & Mehl, 2015). Without pretending to be exhaustive, I posited that SJTs and assessment centre exercises show promise for addressing key research questions related to within-person variability, personality disorders, trait-behaviour links, and personality expression and perception. I welcome other potential avenues for future research by using these approaches and hope that this article can instigate many fruitful cross-disciplinary investigations and collaborations.

## REFERENCES

- Allik, J., de Vries, R. E., & Realo, A. (2016). Why are moderators of self-other agreement difficult to establish? *Journal of Research in Personality, 63*, 72–83.
- Allport, G. W. (1951). *Personality—A psychological interpretation*. London: Constable.
- Arthur, W. Jr., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology, 56*, 125–153.
- Asendorpf, J. B., & Wilpers, S. (1998). Personality effects on social relationships. *Journal of Personality and Social Psychology, 74*, 1531–1544.
- Back, M. D., Schmukle, S. C., & Egloff, B. (2009). Predicting actual behavior from the explicit and implicit self-concept of personality. *Journal of Personality and Social Psychology, 97*, 533–548.
- Back, M. D., & Nestler, S. (2016). Accuracy of judging personality. In: In J. A. Hall, M., Schmid Mast, & T. V. West. (Eds.) *The social psychology of perceiving others accurately* (pp. 98–124). Cambridge, MA: Cambridge University Press.
- Back, M. D., & Vazire, S. (2015). The social consequences of personality: Six suggestions for future research. *European Journal of Personality, 29*, 296–307.
- Back, M. D., Baumert, A., Denissen, J. J. A., Hartung, F.-M., Penke, L., Schmukle, S. C., Schönbrodt, F. D., ... Wrzus, C. (2011). PERSOC: A unified framework for understanding the dynamic interplay of personality and social relationships. *European Journal of Personality, 25*, 90–107.
- 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.
- Bangerter, A., Roulin, N., & König, C. J. (2012). Personnel selection as a signaling game. *Journal of Applied Psychology, 97*, 719–773.
- Barrick, M. R., & Mount, M. K. (1991). The big 5 personality dimensions and job-performance—A meta analysis. *Personnel Psychology, 44*, 1–26.
- 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.
- Bledow, R., & Frese, M. (2009). A situational judgment test of personal initiative: Towards understanding construct based situational judgment tests. *Personnel Psychology*, 62, 229–258.
- Bleidorn, W., & Ködding, C. (2013). The divided self and psychological (mal)adjustment—A meta-analytic review. *Journal of Research in Personality*, 47, 547–552.
- Block, J., & Block, J. H. (1981). Studying situational dimensions: A grand perspective and some limited empiricism. In D. Magnusson (Ed.), *Toward a psychology of situations: An interactionist perspective* (pp. 85–102). Hillsdale, NJ: Erlbaum.
- Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology*, 86, 599–614.
- Bowler, M. C., & Woehr, D. J. (2006). A meta-analytic evaluation of the impact of dimension and exercise factors on assessment center ratings. *Journal of Applied Psychology*, 91, 1114–1124.
- Brunswik, E. (1952). *The conceptual framework of psychology*. Chicago: University of Chicago Press.
- Campion, M. C., & 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.
- Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgement tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology*, 63, 83–117.
- Christiansen, N. D., & Tett, R. P. (2013). *Handbook of personality at work*. New York, NY: Routledge.
- Christiansen, N. D., Wolcott-Burnam, S., Janovics, J. E., Burns, G. N., & Quirk, S. W. (2005). The good judge revisited: Individual differences in the accuracy of personality judgments. *Human Performance*, 18, 123–149.
- Christiansen, N., Hoffman, B. J., Lievens, F., & Speer, A. (2013). Assessment centers and the measurement of personality. In N. Christiansen, & R. Tett (Eds.), *Handbook of personality at work* (pp. 477–497). New York (NY): Routledge.
- Clifton, A. (2014). Variability in personality expression across contexts: A social network approach. *Journal of Personality*, 82, 103–115.
- Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, 136, 1092–1122.
- Cucina, J. M., Chihwei, S., 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.
- Dalal, R. S., Meyer, R. D., Bradshaw, R. P., Green, J. P., Kelly, E. D., & Zhu, M. (2015). Personality strength and situational influences on behavior: A conceptual review and research agenda. *Journal of Management*, 41, 261–287.
- De Kock, F., Lievens, F., & Born, M. (2015). An in-depth look at dispositional reasoning and interviewer accuracy. *Human Performance*, 28, 199–221.
- Dunlop, W. L. (2015). Contextualized personality, beyond traits. *European Journal of Personality*, 29, 310–325.
- Ebner-Priemer, U. W., Eid, M., Kleindienst, N., Stabenow, S., & Trull, T. J. (2009). Analytic strategies for understanding affective (in)stability and other dynamic processes in psychopathology. *Journal of Abnormal Psychology*, 118, 195–202.
- Erickson, T. E., 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.
- Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, 80, 1011–1027.
- 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.
- Fleeson, W., & Jayawickreme, E. (2015). Whole trait theory. *Journal of Research in Personality*, 56, 82–92.
- Fleeson, W., & Law, M. K. (2015). Trait manifestations as density distributions: The role of actors, situations, and observers in explaining stability and variability. *Journal of Personality and Social Psychology*, 109, 1090–1104.
- Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception*. San Diego: Academic Press.
- Funder, D. C., Furr, R. M., & Colvin, C. R. (2000). The riverside behavioral Q-sort: A tool for the description of social behavior. *Journal of Personality*, 68, 451–489.
- Furnham, A., & Jaspars, J. (1983). Interactionism in psychology: A critical analysis of the situation–response inventories. *Personality and Individual Differences*, 4, 627–644.
- Furr, R. M. (2009). Personality psychology as a truly behavioral science. *European Journal of Personality*, 23, 369–401.
- Gatewood, R., Field, H. S., & Barrick, M. (2015). *Human resource selection* (8th ed.). Boston: Cengage Learning.
- Geukes, K., Nestler, S., Hutteman, R., Küfner, A. C. P., & Back, M. D. (in press). Trait personality and state variability: Predicting individual differences in within- and cross-context fluctuations in affect, self-evaluations, and behavior in everyday life. *Journal of Research in Personality*.
- Gibbons, A. M., & Rupp, D. E. (2009). Dimension consistency as an individual difference: A new (old) perspective on the assessment center construct validity debate. *Journal of Management*, 35, 1154–1180.
- Giddens, A. (1991). *Modernity and self-identity: Self and society in the late modern age*. Cambridge, UK: Polity Press.
- Goffman, E. (1959). *The presentation of self in everyday life*. New York: Anchor.
- Gosling, S. D., John, O. P., Craik, K. H., & Robins, R. W. (1998). Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. *Journal of Personality and Social Psychology*, 74, 1337–1349.
- Haselton, M. G., & Funder, D. C. (2006). The evolution of accuracy and bias in social judgment. In M. Schaller, J. A. Simpson, & D. T. Kenrick (Eds.), *Evolution and social psychology* (pp. 15–38). New York, NY: Psychology Press.
- Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, 57, 639–683.
- Herde, C.N., & Lievens, F. (2016, April). *The added value of speed assessment for assessing leadership*: In P. V. Ingold, & B. J. Hoffman (Chair), AC, Quo Vadis? Novel assessment center design and validation strategies. Symposium conducted at the Annual Conference of the Society for Industrial and Organizational Psychology, Anaheim, CA, USA.
- Hirschmüller, S., Egloff, B., Schmukle, S. C., Nestler, S., & Back, M. D. (2015). Accurate judgments of neuroticism at zero acquaintance: A question of relevance. *Journal of Personality*, 83, 221–228.
- Hoffman, B. J., Kennedy, C. L., LoPilato, A. C., Monahan, E. L., & Lance, C. E. (2015). A review of the content, criterion-related, and construct-related validity of assessment center exercises. *Journal of Applied Psychology*, 100, 1143–1168.
- Hofstee, W. K. B. (1994). Who should own the definition of personality? *European Journal of Personality*, 8, 149–162.
- Hogan, R., & Roberts, B. W. (2000). A socio analytic perspective on person environment interaction. In W. B. Walsh, K. H. Craik, & R. H. Price (Eds.), *Person-environment psychology: New directions and perspectives* (pp. 1–23). Mahwah, NJ: Erlbaum.
- Hogan, R., & Shelton, D. (1998). A socioanalytic perspective on job performance. *Human Performance*, 11, 129–144.
- Huang, J. L., Ryan, A. M., Zabel, K. L., & Palmer, A. (2014). Personality and adaptive performance at work: A meta-analytic investigation. *Journal of Applied Psychology*, 99, 162–179.



- Human, L. J., & Biesanz, J. C. (2013). Targeting the good target: An integrative review of the characteristics and consequences of being accurately perceived. *Personality and Social Psychology Review, 17*, 248–272.
- Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology, 85*, 869–879.
- Ickes, W., Snyder, M., & Garcia, S. (1997). Personality influences on the choice of situations. In R. Hogan, & J. A. Johnson (Eds.), *Handbook of personality psychology* (pp. 165–195). San Diego, CA: Academic Press, Inc.
- James, L. R. (1998). Measurement of personality via conditional reasoning. *Organizational Research Methods, 1*, 131–163.
- Johnson, W. (2015). Editorial. *European Journal of Personality, 29*, 309.
- Judge, T. A., & Zapata, C. P. (2015). The person-situation debate revisited: Effect of situation strength and trait activation on the validity of the Big Five traits in predicting job performance. *Academy of Management Journal, 58*, 1–31.
- Judge, T. A., Simon, L. S., Hurst, C., & Kelley, K. (2014). What I experienced yesterday is who I am today: Relationship of work motivations and behaviors to within-individual variation in the five-factor model of personality. *Journal of Applied Psychology, 99*, 199–221.
- Kanning, U. P., Grewe, K., Hollenberg, S., & Hadouch, M. (2006). From the subjects' point of view: Reactions to different types of situational judgment items. *European Journal of Psychological Assessment, 22*, 168–176.
- Kenny, D. A. (2004). PERSON: A general model of interpersonal perception. *Personality and Social Psychology Review, 8*, 265–280.
- Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1993). The 'Trier Social Stress Test'—A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology, 28*, 76–81.
- Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological Medicine, 42*, 1879–1890.
- 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.
- Labrador, J. R., Christiansen, N. D., & Burns, G. N. (2006, April). *Measuring personality using situational judgment tests*. Paper presented at the 21st annual conference of the Society for Industrial and Organizational Psychology, Dallas, TX.
- Leckelt, M., Küfner, A. C. P., Nestler, S., & Back, M. D. (2015). Behavioral processes underlying the decline of narcissists' popularity over time. *Journal of Personality and Social Psychology, 109*, 856–871.
- Leikas, S., Lönnqvist, J.-E., & Verkasalo, M. (2012). Persons, situations, and behaviors: Consistency and variability of different behaviors in four interpersonal situations. *Journal of Personality and Social Psychology, 103*, 1007–1022.
- Letzring, T. D. (2008). The good judge of personality: Characteristics, behaviors, and observer accuracy. *Journal of Research in Personality, 42*, 914–932.
- Levy, D. A., Collins, B. E., & Nail, P. R. (1998). A new model of interpersonal influence characteristics. *Journal of Social Behavior & Personality, 13*, 715–733.
- Lievens, F., & Motowidlo, S. J. (2015). Situational judgment tests: From measures of situational judgment to measures of general domain knowledge. *Industrial and Organizational Psychology – Perspectives on Science and Practice, 9*, 3–22.
- 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.
- Lievens, F., & Sackett, P. (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.
- Lievens, F., & Sackett, P. R. (2017). The effects of predictor method factors on selection outcomes: A modular approach to personnel selection procedures. *Journal of Applied Psychology, 102*, 43–66.
- Lievens, F., Chasteen, C. S., Day, E. A., & Christiansen, N. D. (2006). Large-scale investigation of the role of trait activation theory for understanding assessment center convergent and discriminant validity. *Journal of Applied Psychology, 91*, 247–258.
- Lievens, F., De Corte, W., & Schollaert, E. (2008). A closer look at the frame-of-reference effect in personality scale scores and validity. *Journal of Applied Psychology, 93*, 268–279.
- Lievens, F., De Corte, W., & Westerveld, L. (2015a). Understanding the building blocks of selection procedures: Effects of response fidelity on performance and validity. *Journal of Management, 41*, 1604–1627.
- Lievens, F., De Fruyt, F., & Van Dam, K. (2001). Assessors' use of personality traits in descriptions of assessment centre candidates: A five-factor model perspective. *Journal of Occupational and Organizational Psychology, 74*, 623–636.
- Lievens, F., Lang, J., Vandevijver, M., Corstjens, J., De Fruyt, F., & Bledow, R. (in press). A closer look at the assessment and predictive power of people's intra-individual variability across situations. *Journal of Applied Psychology*.
- Lievens, F., Schollaert, E., & Keen, G. (2015b). The interplay of elicitation and evaluation of trait-expressive behavior: Evidence in assessment center exercises. *Journal of Applied Psychology, 100*, 1169–1188.
- Lievens, F., Tett, R. P., & Schleicher, D. J. (2009). Assessment centers at the crossroads: Toward a reconceptualization of assessment center exercises. In J. J. Martocchio, & H. Liao (Eds.), *Research in personnel and human resources management* (pp. 99–152). Bingley: JAI Press.
- MacCann, C., & Roberts, R. D. (2008). New paradigms for assessing emotional intelligence: Theory and data. *Emotion, 8*, 540–551.
- McCarthy, J. M., Van Iddekinge, C. H., Lievens, F., Kung, M. C., Sinar, E. F., & Campion, M. A. (2013). Do candidate reactions relate to job performance or affect criterion related validity? A multistudy investigation of relations among reactions, selection test scores, and job performance. *Journal of Applied Psychology, 98*, 701–719.
- McCrae, R. R., & Costa, P. T. (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* (pp. 51–87). New York: Guilford Press.
- McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. L. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology, 60*, 63–91.
- McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Predicting job performance using situational judgment tests: A clarification of the literature. *Journal of Applied Psychology, 86*, 730–740.
- Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology, 90*, 862–877.
- Meriac, J. P., Hoffman, B. J., Woehr, D. J., & Fleisher, M. S. (2008). Further evidence for the validity of assessment center dimensions: A meta-analysis of the incremental criterion-related validity of dimension ratings. *Journal of Applied Psychology, 93*, 1042–1052.
- Meyer, R. D., Dalal, R. S., & Hermida, R. (2010). A review and synthesis of situational strength in the organizational sciences. *Journal of Management, 36*, 121–140.
- Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review, 80*, 252–283.
- Mischel, W. (2009). From personality and assessment (1968) to personality science (2009). *Journal of Research in Personality, 43*, 282–290.

- 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.
- Moskowitz, D. S., & Zuroff, D. C. (2004). Flux, pulse, and spin: Dynamic additions to the personality lexicon. *Journal of Personality and Social Psychology*, *86*, 880–893.
- Motowidlo, S. J., & Beier, M. E. (2010). Differentiating specific job knowledge from implicit trait policies in procedural knowledge measured by a SJT. *Journal of Applied Psychology*, *95*, 321–333.
- Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006a). Implicit policies about relations between personality traits and behavioral effectiveness in situational judgment items. *Journal of Applied Psychology*, *91*, 749–761.
- Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006b). A theoretical basis for situational judgment tests. In J. A. Weekley, & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 57–82). New Jersey: Lawrence Erlbaum.
- Mueller-Hanson, R., Heggstad, E. D., & Thornton, G. C. (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology*, *88*, 348–355.
- Murphy, N. A., & Hall, J. A. (2011). Intelligence and interpersonal sensitivity: A meta-analysis. *Intelligence*, *39*, 54–63.
- Murray, H. (1938). *Explorations in personality*. New York: Oxford University Press.
- Mussel, P., Gatzka, T., & Hewig, J. (2016). Situational judgment tests as an alternative measure for personality assessment. *European Journal of Psychological Assessment*, Online Publication, August 3. <http://dx.doi.org/10.1027/1015-5759/a000346>
- Nestler, S., Egloff, B., Küfner, A. C. P., & Back, M. D. (2012). An integrative lens model approach to bias and accuracy in human inferences: Hindsight effects and knowledge updating in personality judgments. *Journal of Personality and Social Psychology*, *103*, 698–717.
- Oh, I. S., Wang, G., & Mount, M. K. (2011). Validity of observer ratings of the five-factor model of personality traits: A meta-analysis. *Journal of Applied Psychology*, *96*, 762–773.
- Oliver, T., Hausdorf, P., Lievens, F., & Conlon, P. (2016). Interpersonal dynamics in assessment center exercises: Effects of role player portrayed disposition. *Journal of Management*, *42*, 1992–2017.
- Olson-Buchanan, J. B., Drasgow, F., Moberg, P. J., Mead, A. D., Keenan, P. A., & Donovan, M. (1998). Conflict resolution skills assessment: A model-based, multi-media approach. *Personnel Psychology*, *51*, 1–24.
- Parrigon, S., Sang, S. E., Tay, L., & Wang, T. (2017). CAPTIONing the situation: A lexically-derived taxonomy of psychological situation characteristics. *Journal of Personality and Social Psychology*, *112*, 642–681.
- Pulakos, E. D., Schmitt, N., Whitney, D., & Smith, M. (1996). Individual differences in interviewer ratings: The impact of standardization, consensus discussion, and sampling error on the validity of a structured interview. *Personnel Psychology*, *49*, 85–102.
- Ragsdale, J. M., Rahael, J. A., Christiansen, N. D., Frost, C., & Burns, G. (2013). Using content analysis in the assessment of personality. In N. D. Christiansen, & R. Tett (Eds.), *Handbook of personality at work* (pp. 498–522). New York (NY): Routledge.
- Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., ... Funder, D. C. (2014). The situational eight DIAMONDS: A taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology*, *107*, 677–718.
- 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.
- Roberts, B. W. (2007). Contextualizing personality psychology. *Journal of Personality*, *75*, 1071–1082.
- Robins, R. W., & Beer, J. S. (2001). Positive illusions about the self: Short-term benefits and long-term costs. *Journal of Personality and Social Psychology*, *80*, 340–352.
- Roch, S. G., Woehr, D. J., Mishra, V., & Kieszczynska, U. (2012). Rater training revisited: An updated meta-analytic review of frame-of-reference training. *Journal of Occupational and Organizational Psychology*, *85*, 370–395.
- 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.
- Sackett, P. R., & Lievens, F. (2008). Personnel selection. *Annual Review of Psychology*, *59*, 419–450.
- Sackett, P. R., Lievens, F., Van Iddekinge, C., & Kuncel, N. (2017). Individual differences and their measurement: A review of 100 years of research. *Journal of Applied Psychology*, *102*, 254–273.
- Schlenker, B. R., & Leary, M. R. (1982). Social anxiety and self-presentation: A conceptualization model. *Psychological Bulletin*, *92*, 641–669.
- Schollaert, E., & Lievens, F. (2012). Building situational stimuli in assessment center exercises: Do specific exercise instructions and role-player prompts increase the observability of behavior. *Human Performance*, *25*, 255–271.
- Shaffer, J. A., & Postlethwaite, B. E. (2012). A matter of context: A meta-analytic investigation of the relative validity of contextualized and noncontextualized personality measures. *Personnel Psychology*, *65*, 445–494.
- Sherman, R. A., Nave, C. S., & Funder, D. C. (2010). Situational similarity and personality predict behavioral consistency. *Journal of Personality and Social Psychology*, *99*, 330–343.
- Sherman, R. A., Rauthmann, J. F., Brown, N. A., Serfass, D. S., & Jones, A. B. (2015). The independent effects of personality and situations on real-time expressions of behavior and emotion. *Journal of Personality and Social Psychology*, *109*, 872–888.
- Simms, L. J., Goldberg, L. R., Roberts, J. E., Watson, D., Welte, J., & Rotterman, J. H. (2011). Computerized adaptive assessment of personality disorder: Introducing the CAT-PD project. *Journal of Personality Assessment*, *93*, 380–389.
- Slatcher, R. B., & Vazire, S. (2009). Effects of global and contextualized personality on relationship satisfaction. *Journal of Research in Personality*, *43*, 624–633.
- Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, *30*, 526–537.
- Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology*, *88*, 500–517.
- 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.
- Thornton, G. C., & Rupp, D. E. (2006). *Assessment centers in human resource management: Strategies for prediction, diagnosis, and development*. Mahwah, NJ: Lawrence Erlbaum.
- Thornton, G. C. III, & Cleveland, J. N. (1990). Developing managerial talent through simulation. *American Psychologist*, *45*, 190–199.
- Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review*, *93*, 239–257.
- Trull, T. J., Solhan, M. B., Tragesser, S. L., Jahng, S., Wood, P. K., Piasecki, T. M., & Watson, D. (2008). Affective instability: Measuring a core feature of borderline personality disorder with ecological momentary assessment. *Journal of Abnormal Psychology*, *117*, 647–661.
- Van Iddekinge, C. H., Raymark, P. H., & Roth, P. L. (2005). Assessing personality with a structured employment interview: Construct-related validity and susceptibility to response inflation. *Journal of Applied Psychology*, *90*, 536–552.

- Van Iddekinge, C. H., Sager, C. E., Burnfield, J. L., & Heffner, T. S. (2006). The variability of criterion-related validity estimates among interviewers and interview panels. *International Journal of Selection and Assessment, 14*, 193–205.
- Vazire, S. (2010). Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology, 98*, 281–300.
- Vazire, S., & Mehl, M. R. (2008). Knowing me, knowing you: The accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *Journal of Personality and Social Psychology, 95*, 1202–1216.
- Wang, L., & Grimm, K. J. (2012). Investigating reliabilities of intraindividual variability indicators. *Multivariate Behavior Research, 47*, 1–31.
- 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.
- Weis, S., & Süß, H.-M. (2005). Social intelligence—A review and critical discussion of measurement concepts. In R. Schulze, & R. D. Roberts (Eds.), *Emotional intelligence—An international handbook* (pp. 203–230). Cambridge: Hogrefe & Huber Publishing.
- Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance-appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology, 67*, 189–205.
- Wrzus, C., & Mehl, M. (2015). Lab and/or field? Measuring personality processes and their social consequences. *European Journal of Personality, 29*, 250–271.

#### APPENDIX: EXAMPLE OF THE COMPUTATION OF IMPLICIT TRAIT POLICIES RELATED TO AGREEABLENESS AND EXTRAVERSION

As shown in the table below, a target person's ITP associated with agreeableness is computed by correlating the values related to agreeableness across items in the two columns displayed in *italic*. A target person's ITP associated with extraversion is computed by correlating the values related to extraversion across items in the two columns displayed in **bold**.

	Trait level of response option (validated via subject matter experts; 0 = low and 1 = high)	Target person's rating of effectiveness of response option (from 1 = highly ineffective to 7 = highly effective)
SJT item 1: Option A ( <i>low agreeableness</i> )	<i>0</i>	<i>3</i>
: Option B ( <i>high agreeableness</i> )	<i>1</i>	<i>6</i>
: Option C ( <i>low agreeableness</i> )	<i>0</i>	<i>4</i>
: Option D ( <i>high agreeableness</i> )	<i>1</i>	<i>4</i>
SJT item 2: Option A ( <i>low extraversion</i> )	<b>0</b>	<b>2</b>
: Option B ( <i>high extraversion</i> )	<b>1</b>	<b>7</b>
: Option C ( <i>low extraversion</i> )	<b>0</b>	<b>2</b>
: Option D ( <i>high extraversion</i> )	<b>1</b>	<b>5</b>
SJT item 3: Option A ( <i>low agreeableness</i> )	<i>0</i>	<i>3</i>
: Option B ( <i>low extraversion</i> )	<b>0</b>	<b>1</b>
: Option C ( <i>high agreeableness</i> )	<i>1</i>	<i>1</i>
: Option D ( <i>high extraversion</i> )	<b>1</b>	<b>5</b>
Etc.	Etc.	Etc.
SJT item 30: Option A ( <i>low agreeableness</i> )	<i>0</i>	<i>3</i>
: Option B ( <i>low agreeableness</i> )	<i>0</i>	<i>1</i>
: Option C ( <i>high agreeableness</i> )	<i>1</i>	<i>1</i>
: Option D ( <i>high extraversion</i> )	<b>1</b>	<b>7</b>