OPEN PEER COMMENTARY

Employing Situational Simulations to Understand Person-Situation Transaction Processes

ANNA BAUMERT[1,2] and GABRIELA BLUM[3]

[1]*Max-Planck Institute for Research on Collective Goods, Bonn, Germany*
[2]*Technical University Munich, School of Education, Munich, Germany*
[3]*University of Koblenz–Landau, Landau, Germany*
baumert@coll.mpg.de

Abstract: Drawing on developments in personnel selection, Lievens proposed using Situational Judgment Tests and Assessment Centre exercises for personality research. These instruments simulate situations to elicit (typical or maximal) responses. Building on Lievens's suggestions, we elaborate on how such simulations can be used to disentangle processes of person-situation transactions. However, we caution against potential lack of validity of this measurement approach. Behavioural domains may vary in response validity in simulations, and we discuss some potential moderators of validity. Copyright © 2017 European Association of Personality Psychology

Lievens proposed that situation simulation as used in Situational Judgment Tests (SJTs) and Assessment Centre exercises (ACs) can serve to study between- and within-person variability in behaviour more generally. We elaborate on how this approach can be refined to disentangle different kinds of person-situation transactions (Caspi, 1998) and the processes involved (Blum & Schmitt, 2017).

Processes of reactive transactions could be disentangled by varying situational cues involved in simulated situations. By asking participants about their perceptions of these situations, we learn about individual differences in tendencies to attend to certain kinds of stimuli and interpret them in certain ways. Moreover, simulated situations can be disambiguated (particularly in written descriptions) by specifying situation characteristics (i.e. terms that describe situation attributions, such as 'dangerous' or 'stressful'). Assessing participants' reactions then isolates individual differences in reactivity from perceptual tendencies. Employing open-ended responses or having participants enact their reactions to situations, as Lievens proposed, further reveals individual differences in behavioural repertoires.

Processes of proactive, manipulative, and evocative transactions could be disentangled by eliciting various kinds of behavioural responses to simulated situations, with the decisions determining what further situations will be encountered, as Lievens suggests. (Think of so-called Gamebooks, i.e. branching-plot novels that allow the reader to choose his or her own adventure.) Choices to enter or avoid certain kinds of situations (proactive) or to shape situations actively (e.g. by acquiring objects or skills; manipulative), as well as behavioural decisions within social interactions that elicit interaction-partner behaviours (evocative) could be assessed separately. Employing such adaptive questionnaires could then simulate how individual differences in behavioural tendencies might be potentiated (or diminished) by complex patterns of transactions.

Assessment of individual differences in these processes could be helpful in explaining individual differences in behaviour better. In combination with experimental approaches that directly manipulate the respective processes, causes of individual differences in behaviour can be revealed. Similar approaches have been used in the past (e.g. Lawson & MacLeod, 1999). Extending them to cover broad ranges of behavioural domains systematically (e.g. situational cues and perceptions pertaining to all DIAMONDs, Rauthman et al., 2014) will facilitate determining whether person-situation transaction processes are indeed generic mechanisms involved in many behavioural domains, or whether some processes might be involved in some domains, but not others.

Contextualized assessment of individual differences that captures each of the named processes in a variety of contexts (e.g., work, health, and family) might also allow better predictions of future behaviour or behavioural outcomes. In accordance with the well-established principle of correspondence or symmetry (Ajzen & Fishbein, 1977; Brunswik, 1956; Epstein, 1979; Hofmann, Gschwendner, Nosek, & Schmitt, 2005), Lievens reported that contextualized measures have greater predictive power regarding contextualized outcomes than non-contextualized measures. However, even for predictions of highly generalized outcomes, such as life satisfaction, aggregation of contextualized items across behavioural and processing tendencies and across contexts might be preferable to typical personality questionnaire items. The latter require participants to aggregate their behavioural and processing tendencies across a broad range of situations and time (e.g. 'I remain unemotional even in situations where most people get very sentimental', Ashton & Lee, 2009). It seems likely that responses to these items

are subject to biases in memory retrieval and utilization, such that they selectively reflect some behavioural and processing tendencies and some contexts more than others. Thus, employment of contextualized items from a broad range of contexts might capture a targeted construct in its full range.

MODERATORS OF VALIDITY

Social and personality psychology both have traditions of employing self-reported reactions to written descriptions of hypothetical situations (i.e. vignettes, comparable in relevant regards with SJTs; Baumeister, Vohs, & Funder, 2007). However, validity of these assessments has been questioned. Even under high degrees of correspondence between simulated situations and actually experienced situations, self-reported predictions of own behaviour and behavioural observations have been found to differ (e.g., Baumert, Halmburger, & Schmitt, 2013; LaPiere, 1934; Wicker, 1969), both in mean levels and in rank-orders.

On one hand, low fidelity on the side of the situation simulation can be blamed. In particular, written descriptions present information in linear and disambiguated manners that are unlikely to match how real situations unfold. For this reason, use of auditory or video material seems favourable. This is also true for ACs if they involve written background information or instructions for role plays (e.g. 'imagine, at a working day that is already packed with work, two colleagues come to your office to search your help concerning a conflict …').

On the other hand, individuals might not be capable or motivated to report their behavioural inclinations correctly in particular situations. Behavioural domains might vary in how well people can predict their own responses. We hypothesize that in cases of high social desirability of particular responses, people might be motivated to report their behavioural inclinations more favourably and some people more than others (see Lievens's review of findings on faking). We further hypothesize that people might be less able correctly to predict their own reactions in cases of limited experience with relevant situations (e.g. emergencies). Finally, behavioural domains might vary in degree to which conflicting processes shape behaviour. The more complex the processes that shape behaviour, the more likely it seems that people inaccurately predict their own behaviour. Tests of these hypotheses will help to make possible situation simulations in domains (and for people) with high validity.

Using Situational Judgment Tests and Assessment Centres in Personality Psychology: Three Suggestions

SIMON M. BREIL, KATHARINA GEUKES and MITJA D. BACK

University of Münster, Münster, Germany
simon.breil@uni-muenster.de

Abstract: Lievens called for integration to personality research of behavioural assessment procedures developed in selection research. We agree and describe three exemplary fields of application: better understanding of behavioural preferences and expectancies, investigation of situational characteristics that drive personality expression, and analysis of individual differences in intra-individual behavioural variability within and across contexts. When carefully considering differences between maximum and typical performance as well as between reported and actual behaviour, selection and personality researchers can effectively work together to understand the personality triad of persons, situations, and behaviours better. Copyright © 2017 European Association of Personality Psychology

Selection and personality researchers are both interested in predicting individual differences in real-life behaviours in circumscribed situations; however, there has been little cross-talk between these disciplines. Lievens described two classes of assessment procedures, Situational Judgment Tests (SJTs) and Assessment Centre exercises (ACs), that have been mostly developed independently of personality research. Adopting these techniques more widely in personality psychology indeed provides unique opportunities for better understanding the personality triad (persons, situations, and behaviours; Bem, 1983; Funder, 2006).

We discuss three ways of integrating selection research into personality research that we regard as particularly promising: (1) better understanding of behavioural preferences and expectancies through implicit trait policies (ITP) as assessed in SJTs, (2) investigation of situational characteristics that drive personality expression via ACs, and (3) analysis of individual differences in intra-individual behavioural variability within and across contexts via ACs and SJTs. For all these domains of integration, one should be aware of the basic distinction between maximum performance (e.g. behaving as attentively as one can) and typical performance (e.g. behaving as attentively as one usually does; cf. Sackett, Zedeck, & Fogli, 1988). The situational factors that evoke differences in maximum performance in selection contexts do not necessarily need to be the same as those that evoke differences in typical performance in many of the non-selection contexts personality research typically targets. Furthermore, there is an important distinction between behavioural reports as assessed with SJTs (that measure some sort of behavioural preference or expectancy) and direct behavioural observations as assessed with ACs (that measure actual behaviour). Both assessment approaches can be relevant to personality psychology, but

they should not be treated as if they represent the identical phenomenon (e.g. Back & Egloff, 2009; Back, Schmukle, & Egloff, 2009; Vazire & Mehl, 2008).

First, ITPs in selection research are typically measured for each trait as the within-person correlations between the expected effectiveness of a number of behavioural alternatives (what should you do?) and expert-ratings on whether behaviours represent high or low values on specific traits. ITPs, thus, regard people's beliefs about the relative effectiveness of maximum performance in specified behavioural domains (e.g. how much agreeable behaviour helps to make meetings with challenging co-workers effective, cf. Motowidlo, Hooper, & Jackson, 2006). When translating ITPs to the domains of typical behaviour (what would you do?) in personality psychology, ITPs might, for example, be adopted to represent alternative measures of people's motives and goals. Items could refer to ambiguous social situations (e.g. an awkward group interaction in which some are consuming resources at the expense of others), and rated behavioural response options could represent different strategies in goal achievement that are classified (by expert ratings or empirical findings) as communal, 'getting along' versus agentic, and 'getting ahead' behaviours. By experimenting with different instructions, one might disentangle further individual differences involved in the construction of behavioural intentions (e.g. 'Ideal self': What one prefers to do; 'Ought self': What one thinks one should do; Self-efficacy expectations: What one thinks one is able to do.).

Second, AC exercises can be used to operationalize situational characteristics that reliably highlight specific traits and, thus, evoke behavioural differences in trait expression. These AC-like tasks allow creation of a large set of situations that systematically vary regarding both situational context and interaction partner variables. Regarding context variables, one might, for example, vary the number of persons present, the degree of social interaction or social stress, and task complexity. These characteristics can then be correlated with the observed actual variability between individuals in trait-related behaviours across situations to identify the most important aspects for trait-activation (e.g. the degree of social stress might be correlated with between-person variability in nervous behaviour; e.g. Hirschmüller, Egloff, Schmukle, Nestler, & Back, 2015). Similarly, regarding interaction partner characteristics, role player prompts (cf. Schollaert & Lievens, 2011, also see Leikas, Lönnqvist, & Verkasalo, 2012 for similar procedures) can be used to evoke specifically targeted behavioural differences. For example, one might evoke differences in (dis)agreeable behaviour by asking role players to show challenging behaviour towards a participant. Role players may even actively vary trait-relevance within a given situation to assess multiple behaviours in one sitting (e.g. first act shy to evoke differences in extraverted behaviour, then act challenging for agreeable behaviour). We believe that the systematic variation of naturally occurring and psychologically relevant contextual features in combination with use of role players who produce varying and adaptive situational cues represents a fruitful endeavour for behavioural personality research.

Third, combining the above, SJTs and ACs together can be used to understand individual differences in intra-individual behavioural variability better—both on a conceptual level (i.e. how much people differ in behavioural self-concept; SJT) and an actual behavioural level (i.e. how much people actually differ in expressed behaviour; AC). According to the Within and Across Context framework (Geukes, Nestler, Hutteman, Küfner, & Back, 2017), overall variability across measurement occasions should be parsed into variability within context and variability across context. Applying this concept to ITPs measured via SJTs, one could use the variability of trait-specific ITP correlations across similar situational descriptions as a measure of within-context variability in behavioural goals or strategies, whereas variability across differing situational descriptions provides a measure of cross-context variability. Within-context variability in an AC can be accessed via measuring behaviour multiple times within one situation (e.g. behaviour in the first/second/third minute) and cross-context variability with average within-situation assessments across differing situations. This approach provides a promising method to measure actual behavioural variability in controlled settings.

In sum, we fully support Lievens's call for more crosstalk. When considering differences between maximum and typical performance as well as between reported and actual behaviour, assessment techniques developed in selection research can be fruitfully integrated into personality psychology. This applies in particular to understanding of individual differences in striving for expression and variability of actual behaviours.

Putting Situations into Personality Assessments: Problems and Potential

NEIL D. CHRISTIANSEN[1] and ANDREW B. SPEER[2]

[1]*Central Michigan University, Mount Pleasant, MI, USA*
[2]*Wayne State University, Detroit, MI, USA*
ookook123@gmail.com

Abstract: Contextualizing personality assessments to involve observing behaviour within situations where specific behaviours and outcomes are to be observed has been shown to enhance validity. We therefore agree with Lievens's endeavour to demonstrate how contextualizing assessments in personality psychology can be expanded outside the workplace. However, this can also limit predictions when there are even small differences in situational demands or inaccuracies in specifying those demands. Additional methods of assessing personality through observer ratings at work and biographical questionnaires are discussed in terms of advantages for explaining behaviour. Copyright © 2017 European Association of Personality Psychology

Conventional wisdom in assessing personality has involved making inferences about traits from observing behaviour (often with self-reported personality inventories) in general or across ranges of situations. Although this has been shown to broadly predict behavior, Lievens correctly points out that our ability to predict behavior is enhanced when it is observed in situations that are similar to those where behavior will be explained (e.g. Sherman, Nave, & Funder, 2010; Shaffer & Postlethwaite, 2012). As such, incorporating information about situations into personality assessments has a great deal to offer personality psychology. However, we note some potential pitfalls and additional methods of assessing personality through observations of work behaviour that may inform personality psychologists.

POTENTIAL PITFALLS IN CONTEXTUALIZING PERSONALITY ASSESSMENTS

As Lievens noted, the key to contextualizing personality assessments to involve relevant situations is identifying what the psychological demands are. However, correctly predicting trait activation potential in situations has proven to be somewhat difficult. For example, in the assessment centre exercises that Lievens discussed, psychologists were concerned that the convergence of ratings on the trait dimensions across exercises was often worse than the correlations of ratings from different traits within exercises (Sackett & Dreher, 1982). Understandably, this led researchers to question the construct validity of the dimension ratings. Studies have also shown better convergence when the exercise situations were similar (Highhouse & Harris, 1993) and especially when they have cues for activating similar traits (Lievens, Chasteen, Day, & Christiansen, 2006). However, when we examined whether composites from similar exercises predicted job performance (a complex outcome involving behaviour across different situations), they fared worse than composites from exercises with different demands (Speer, Christiansen, Goffin, & Goff, 2014). It is therefore possible to contextualize personality assessments based on too narrow a range of demands ('situational bandwidth') when not properly aligned with the behaviours of interest.

In addition, no systematic way of categorizing situations exists, so the most common method of identifying trait activation potential is to ask experts about the behaviours most commonly observed. Recent research has made advances in this area by developing a framework of job characteristics intended to improve the validity of personality measures (Burrus & Way, 2017); however, it has also been shown that experts making ratings such as those used in this framework are not very accurate at identifying relevant traits (Coaster and Christiansen, 2011). In our research to develop a behavioural observation assessment for work simulations (Speer, Christiansen, & Honts, 2015), after viewing the videos of candidates, we believed that the situations put considerable pressure to be assertive and stay calm. Based on our knowledge of personality psychology, we expected Extraversion and Emotional Stability to show the strongest evidence of validity. However, when we collected trait activation potential ratings from assessors, Emotional Stability had the lowest mean of the Five-Factor Model (FFM) traits. The good news was that the ranking of the trait validity from the assessment mirrored the ranking of the trait activation potential by the experts; the bad news was how easy it would have been to be incorrect.

Table 1. Trait-relevant multisource feedback observer ratings from the workplace

Five-factor model dimension	Multiple R for relevant observer dimensions	Salient observer Dimensions from multisource feedback	Example multisource Feedback items
Extraversion	.30	Influences others	Negotiates persuasively
		Speaks effectively	Speaks with enthusiasm and expressiveness
		Drive for results	Displays a high energy level
Agreeableness	.38	Fosters teamwork	Values the contributions of all team members
		Builds relationships	Can be approached easily
		Listens to others	Listens willingly to concerns expressed by others
Conscientiousness	.35	Establishes plans	Develops clear goals and objectives
		Manages execution	Establishes high standards of performance
		Acts with integrity	Lives up to commitments
Emotional stability	.33	Demonstrates adaptability	Involves others in decisions that affect them
		Manages disagreements	Works constructively under stress and pressure
		Displays organizational savvy	Compromises to build relationships with others
Openness to experience	.28	Analyses issues	Understands complex concepts and relationships
		Uses expertise	Keeps up to date on professional/technical issues
		Supports diversity	Understands why people have different beliefs

Note. Multiple R for Relevant Observer Dimensions was based on regressing the self-report FFM scores onto the three salient observer scores from peers and coworkers.

ADDITIONAL OPPORTUNITIES FOR CONTEXTUALIZATION

One approach to contextualizing observations of behaviour at work that has been gaining momentum is to use observer ratings made by coworkers. Research has shown that these ratings may predict work behaviour and job performance better than self-reports (e.g. Connelly, 2013; Kluemper, McLarty, & Bing, 2015). A recent study has also indicated that those ratings explain job performance better than those made by family members or friends, providing support for the importance of contextualization (Kluemper, McLarty, & Bing, 2017). Although unlikely to be useful for external job applicants, it underscored the potential importance for personality researchers to make additional efforts to collect observer ratings from relevant situations rather than relying solely on self-reports. Observer ratings of personality at work may also be useful when considering internal promotions.

In particular, we believe that multisource feedback (MSF) ratings may have untapped potential in this regard. To explore this, we disaggregated the composite ratings in Christiansen and Robie (2011) that had been based on a MSF instrument. For each FFM dimension, we used ratings from the managers' peers and subordinates to identify the most relevant dimensions and behavioural items by correlating ratings with self-reports (Table 1). Convergence was similar to what is typically obtained for different sources of personality ratings. Importantly, the trait-relevant MSF dimension scores were *much* better predictors of supervisors' ratings of job performance ($r = .54$) than were the self-report ratings ($r = .17$). These results underscored advantages of observer ratings from a relevant context.

Personality psychologists have also used questions about biographical data ('biodata' to work psychologists). Example biodata items used to assess Conscientiousness are shown in Table 2. Recent research has demonstrated that focusing on situational demands results in better understanding of cross-situational functionality of empirically keyed biodata. MacLane & Cucina (2015) found that criterion-keying biodata in jobs that required social competence was highly predictive of performance in other jobs that required interpersonal effectiveness but not as predictive for jobs less demanding. Personality psychologists may also find benefits

Table 2. Example biodata items used to assess conscientiousness at work

Item
My attendance record was considered by my past supervisors as
❏ Poor
❏ Fair
❏ Good
❏ Very good
❏ Outstanding
I have made 'to do' lists at work:
❏ Never
❏ Rarely
❏ Sometimes
❏ Often
❏ Very often
I have taken risks on my past jobs:
❏ Much more often than most
❏ Somewhat more often than most
❏ About the same as most
❏ Somewhat less often than most
❏ Much less often than most

from contextualizing questionnaires that ask about past biographical data.

CONCLUSION

We agree wholeheartedly with Lievens about the advantages of contextualizing when observing behaviour for the purpose of assessing personality. However, we do urge some caution, as human judgment is used when gauging the psychological demands. The context may become narrower in bandwidth, or the judges less accurate, than is desired; both of these can result in personality assessments that are less predictive. Even small differences in the demands across situations or slight inaccuracies in specifying those demands can compromise validity. More research into these topics is needed to reduce likelihood of this. We also believe that personality research may benefit by using observer ratings and biographical questionnaires that have been carefully contextualized to include situational demands similar to those impacting the behaviour to be explained.

To Everything There Is a Season: Integrating SJTs and ACs with Traditional Personality Assessments

BRIAN S. CONNELLY[1] and SAMUEL T. MCABEE[2]

[1]University of Toronto, Toronto, Canada
[2]Bowling Green State University, Bowling Green, OH, USA
brian.connelly@utoronto.ca

Abstract: Inclusion of situational judgment tests and assessment centre exercises as assessment methods offers ample opportunities for personality research. These methods provide personality information that partially overlaps with

traditional measures like self-reports but also assess previously untapped elements of personality. We argue that the best way personality researchers can use them is in conjunction with classic assessments. We present an illustrative example for how researchers might incorporate assessment centre ratings with self- and peer-reports of personality building on our Trait-Reputation Identity Model. Copyright © 2017 European Association of Personality Psychology

We appreciate Lievens's thoughtful discussion of how situational judgment tests (SJTs) and assessment centre exercises (ACs) could be used in mainstream personality research. Indeed, selection research has a long and rich history of informing personality theory, from debunking phrenology (Cleeton & Knight, 1924) to Tupes' Air Force cadet selection research entrenching the Five Factor Model (Tupes, 1957; Tupes & Christal, 1958), to landmark meta-analyses and Project A's establishing personality as a predictor of success (e.g. Barrick & Mount, 1991; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990).

Considering SJTs and ACs against the spectrum of personality assessments highlights three focal questions: (1) who should provide the ratings? (2) what situation(s)/ context(s) should the rater consider? and (3) what format do the prompt and response take? Although personality research has taken comfort in the general finding of moderate convergence across many assessment methods (e.g. Connelly & Ones, 2010), voluminous literatures describe how variations in answers to each of these questions can change the meaning of an assessment. For instance, self- and peer-reports of personality are informed by different spheres of knowledge, which shape and reflect how people develop their 'identity' and 'reputation', respectively (Hogan, 1996; Vazire, 2010). Similarly, focusing raters towards specific contexts informs researchers about characteristic adaptations specific to a situation; these unique insights can profile an interpersonal signature that differentiates who a person is across situations (e.g. at home vs at work vs with friends) and over time (e.g. Fleeson, 2007; Fournier, Moskowitz, & Zuroff, 2009). Finally, innovations in scale prompts and response formats can elicit and assess rich processes of behaviours, thoughts, and feelings beyond that available from self- or peer-descriptions. SJTs and ACs (as well as more traditional assessments) each represent particular constellations of answers to these three questions,[1] and the multiple advantages they offer can be cached within their answers therein.

Personality researchers interested in incorporating SJTs and ACs to study particular phenomena (e.g. personality's influence on a romantic partner's relationship satisfaction) will immediately face a dilemma: what should they do with the 'old' measures? Most often, researchers adopting innovative measures have used them either as replacements (i.e. using a new measure instead of an old measure) for or as supplements (e.g. summing scores across scales to produce more precise and representative multi-method measures) to 'tried-and-true' assessment methods. Although adopting replacement or supplement strategies with SJTs and ACs would lead to benefits from using innovative assessments, such strategies can make it difficult to compare findings to those that would be observed from more traditional measures.

In contrast, a third option would use an SJT or AC as a complement to more traditional forms of personality assessment. Specifically, we see inherent value in explicitly studying how the unique lenses of SJTs and/or ACs affect our understanding of the origins and impact of personality. Here, latent variable models provide a valuable tool for separating what a set of measures have in common and have distinct when relating to a criterion (see e.g. McAbee & Connelly, 2016). Thus, if researchers want to use SJTs and ACs to assess something about personality that is innovative and unique from how personality traits have typically been assessed, the best approach is to tease out and set aside SJT and AC variance associated with more traditional measures.

As an illustrative example, suppose a researcher is interested in how others' first impressions of Extraversion predict emergence of leadership, relative to self-views (i.e. Identity) and perceptions of well-acquainted peers (i.e. Reputation) of it. Accordingly, she collects participants' responses on traditional self- and peer-report measures of Extraversion, along with expert ratings of targets' Extraversion from a series of short group-based and individual AC tasks. Subsequently, targets that are unfamiliar with one another participate in a leaderless group discussion task and are rated by group members for their emergence as leaders in the group.

Figure 1 presents a possible representation of this study extending from the Trait-Reputation-Identity Model (McAbee & Connelly, 2016). Here, these various assessments likely have some overlap, reflected in a general 'Trait Extraversion' factor. However, the specific variance captured in each of these assessment modalities (i.e. self-reports, peer-reports, and AC ratings) also captures unique trait insights. For instance, self-reports better assess *less* observable trait-relevant cognitions and feelings (i.e. Identity for Extraversion), whereas peer-reports and ACs provide unique information about *more* observable trait-behaviours (i.e. established reputation and first impressions for Extraversion). The researcher might expect strong, positive effects from both Trait Extraversion and Extraversion first impressions, yet this particular model would allow her to separate variance that is shared among assessments from that which is unique to the AC ratings. We note that similar designs and models could be constructed to (for example) tease apart implicit trait policies in SJTs from explicit trait descriptions in self-reports, or to distinguish generalized from situationally bound personality manifestations in ACs or SJTs.

SJTs, ACs, and personality have traditionally had separate literatures in separate fields, and we applaud Lievens's efforts to highlight their connections and to spur more integrated, innovative assessments of individual differences. We suggest the best innovation will come from designs that adopt purposeful multi-method assessments. Perhaps counterintuitively, if researchers are interested in SJTs or ACs to assess a personality characteristic because of the

---

[1]For a similar discussion of 'modular' assessment choices made across selection instruments, see Lievens and Sackett (2017).
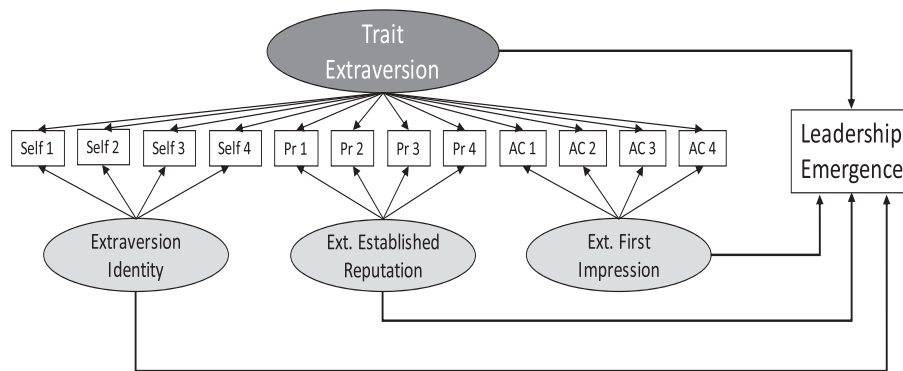
Figure 1.   Hypothetical extension of the Trait-Reputation-Identity model (McAbee & Connelly, 2016) relating self-reports, observer-reports, and assessment centre ratings of agreeableness to performance on a laboratory-based altruism task. 'self' = self-reports; 'Pr' = peer-reports, 'AC' = assessment centre assessments; numbers 1–4 reflect multiple items/assessments of the particular method. [Colour figure can be viewed at wileyonlinelibrary.com]

specific lens the method affords, those researchers can benefit from also including one or more measures of the same construct that lack this particular lens. Such approaches will allow researchers to assess the unique advantages of SJTs, ACs, and many other methods of measuring personality directly.

Applied Personality Assessment: A 'Cronbachian' Perspective

PHILIP J. CORR

City, University of London, London, UK
philip.corr.1@city.ac.uk

Abstract: Lievens's proposal that personality psychology would benefit from using applied tools of assessment—situation judgement tests and assessment centre exercises—was appropriate, especially as these tools focus on real-world criteria in high-stakes situations. Their use would help to integrate (specific) situationally influenced intra-individual differences (variability) and (general) inter-individual differences (diversity). Lievens's proposal also raised a broader issue: each assessment tool yields unique information, and together they have potential to provide a truly comprehensive model of personality based on the 'Cronbachian' perspective that has, so far, not been realized. Copyright © 2017 European Association of Personality Psychology

VARIABILITY AND DIVERSITY

Description and explanation of personality must tackle a peculiarly difficult problem. People share much in common, yet in some ways we are all unique—this is the well-known nomothetic and idiographic distinction. This distinction tends to get lost when we consider complementary aspects of variability and diversity: variability reflects situationally influenced (state) behaviour within individuals and diversity the mean (trait) differences among people. But attempts to provide theoretically coherent integrations of intra-individual variability and inter-individual diversity are fraught with problems. A major one is how the many different sources of variance are best measured, modelled and, importantly, interpreted.

In assessing personality processes, self- and other-report personality questionnaires are valuable. However, measurement of actual behaviour has particular appeal (Furr, 2009), although it is not without its own problems (Corr, 2009). Situational Judgement Tests (SJTs) and Assessment Centre exercise (ACs) have much to offer personality theorists as

well as applied psychologists. The fact that these assessment tools have been developed and tested at the coal face of practical life gives them added credibility.

Lievens's perspective offers the opportunity to integrate specific situationally influenced expressions of personality, potentially covering a wide range of domains (occupational, family, social, leisure and relationships) with personality description and explanation at the broad trait level. While it is true that people react to situational affordances and constraints, each of us possesses stable behavioural patterns (traits) that characterize us. There is nothing contradictory about such a statement—indeed, this was the bedrock of Hans Eysenck's for many decades rather dominant personality theory that postulated the transaction of traits (e.g. Extraversion) with situations (e.g. low vs high arousal), with neither in isolation sufficient to explain behaviour at any one time (Eysenck, 1997). The exciting aspect of Lievens's proposal is that it suggests viable ways to examine these joint effects in a manner that recognizes the importance of both transient states and stable traits: variability and diversity.

COMPLEXITY AND PERPLEXITY

But, on their own, SJTs and ACs do not offer any immediate solution to the many important measurement problems facing personality psychologists; and, indeed, they come with their own limitations: SJTs require the assumption that what people say in reaction to specific (hypothetical) situations reflects how they would behave in situ—but, as we know, self-proclaimed virtue often manifests as behavioural vice. ACs require the assumption that ability to perform well in simulations is tied closely to motivation to do so in real-life situations (e.g. workplace): the difference between 'can do' and 'will do', as well as deliberate faking and such like. However, it might be in such differences in behaviour that insights may be gained into the true, multifaceted, nature of personality.

The complexity of personality psychology forces simplifying modelling choices. This has the consequence of leading to fragmented theory. In the hustle and bustle of scientific life, we trade theoretical comprehensiveness for professional specialization and expertise. In particular, preferred assessment methods are not unrelated to relevant theoretical issues: they constrain the nature of information obtained and thus explanation. For this reason, if for no other, insights into personality processes from related, especially applied, fields should be welcomed. They are badly needed to provide adequate accounts of the complexity, as well as perplexity, of personality psychology.

Along these lines, Poropat and Corr (2015) noted that the development of any integrative model is hindered by the theoretical-epistemological starting point (Popple & Levi, 2000). We search for universals, as seen in personality traits, yet we know that, both as phenomena (Andersen & Chen, 2002) and in assessment (Kenny & West, 2008), traits are socially contextualized. Certainly, this social aspect adds further credibility to Lievens's proposal. What this highlights is that not only are we interested in expression of personality in different situations but often another important source of information comes from raters: their perspectives are central to SJTs and ACs.

A 'CRONBACHIAN' PERSPECTIVE

This discussion may be seen in the context of Cronbach's (1957; Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Cronbach, Rajaratnam, & Gleser, 1963) generalisability theory, which offers the promise of a truly integrative framework for personality assessment and, thus, theory. As detailed by Poropat and Corr (2015), generalisability theory can model the full range of influences in personality assessment, which must include traits, targets, raters, contexts, measurement tools, and temporal factors. As noted by Reynolds et al. (2010) in their detailed discussion of this approach, this goes beyond mere interactionism. It highlights the role of personality judgement as much as expression of personality by the target. There are reasons for supposing that this approach is empirically valuable. For example, such ratings have substantial validity in their own rights (Connelly & Ones, 2010). Although variance in judges' ratings is sometimes denigrated as mere method 'error' or 'bias' (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003; Podsakoff, MacKenzie, & Podsakoff, 2012), it can predict criterion-related behaviours to sufficiently large magnitudes, relative to personality traits alone (Lance, Dawson, Birkelbach, & Hoffman, 2010).

Lievens's proposal for inclusion of SJTs and ACs in personality psychology should be welcomed and is actually rather overdue. However, it can only be part of a broader picture, which encompasses the Cronbachian perspective: where it is assumed that both psychological phenomena and measurement methods contain unique causal variance that is not mere noise. Such 'bias' may well provide unique information on targets (Hoffman & Woehr, 2009; Lance et al., 2010), reflecting 'valid differences in perception' (Borman, 1974, p. 107). According to this Cronbachian perspective, these assessment factors need to be modelled if we are to develop a truly comprehensive personality models. Lievens's proposal can take us a little closer to realizing this goal. It might also get us a little closer to a better understanding of the commonality and uniqueness that characterise personality.

Inspired by Selection Psychology: Yes, Please!

FILIP DE FRUYT, BARBARA DE CLERCQ, LIZE VERBEKE and JASMINE VERGAUWE

Ghent University, Ghent, Belgium
Filip.DeFruyt@ugent.be

Abstract: We elaborate on three points Lievens raised about how assessment methods used in industrial/organizational psychology may advance personality assessment and research. Copyright © 2017 European Association of Personality Psychology

We strongly concur with the position Lievens advocated that the fields of industrial/organizational and personality psychology have much to offer each other (De Fruyt & Salgado, 2003). Lievens outlined several possibilities for how using Situational Judgement Tests (SJT) and Assessment Centre (AC) exercises may fertilize current personality assessment

practices and contemporary personality research. We want to elaborate on two of these suggestions and signal a third application.

First, using the SJT format would indeed be a major step forward in assessing personality pathology and has the potential to form a viable alternative to structured clinical

interviews that are considered as the gold standard for evaluating personality dysfunction. Structured interviews are preferred over self-report measures because they allow asking concrete examples of behaviours, feelings, and cognitions as well as assessing impact or severity of symptoms. Major drawbacks are that interviews are time-consuming, require considerable training to achieve reliable response coding, and assessees select the situations they want to talk about to generate examples of personality manifestations.

The SJT format helps to standardize this idiosyncratic selection of situations, because a common set of situation vignettes representing key dimensions of situational taxonomies can be administered across individuals. Given the strong interpersonal nature and impact of personality disorders, the interpersonal circumplex can serve as a starting point to construct SJT vignettes, enabling systematic evaluation of how people behave across the entire circumplex. SJTs can be self-administered, and the response-coding process can be computerized and standardized. SJTs can bring additional diagnostic information given their flexible answer format. As outlined by Lievens, SJTs can ask people to indicate how they would behave or what they think is the most appropriate behaviour from predefined responses judged by subject matter experts. The first response mode provides information on the typical behaviour that the person would manifest in such situations, whereas the latter type of information reflects more procedural knowledge and one or both of these processes may be impaired in personality-disordered individuals. The distinction between 'not knowing what to do' and 'typically behaving in one or another way' is clinically very useful and implies very different intervention approaches.

Second, Lievens further proposed the SJT format as an economically useful alternative to time-consuming experience-sampling procedures to assess within-person variability, eventually at the cost of ecological validity. Newly developed situational models such as DIAMONDS (Rauthmann et al., 2014) or CAPTION (Parrigon et al., 2016) could help to build situation vignettes and study within-person variability across situation-trait contingencies. The situation vignettes (representing dimensions of the situation models) not only trigger variability in individual's feelings, behaviours, or thoughts but may be also perceived differently by assessees depending on their (disordered) personalities. Such perceptions may mediate situation–trait contingencies and show meaningful between- and within-individual differences. Verbeke and De Clercq (2017) recently developed such an SJT to assess borderline pathology and constructed situational vignettes based upon so-called borderline pathology situational triggers (Miskewicz et al., 2015). In this instrument, situations are written along the DIAMONDS and CAPTION models, whereas the response options reflect various indicators of borderline pathology traits.

Third and finally, there is growing research interest in examining targeted personality change (De Fruyt & Van Leeuwen, 2014; Specht et al., 2014). A key problem in many of these designs is how to evaluate personality change independent of the developing person. Carefully designed ACs, using assessors unfamiliar with the persons' previous trait level and blind to the coached attributes, may help to examine targeted personality change after an intervention or coaching trajectory. The design could include a large number of small assessment exercises (e.g. like in speed assessments) examining an individual's standing on, for example five different traits across some of the exercises, with only one of the traits subjected to intervention. Assessments could eventually be repeated to examine whether change is permanent. Such AC-based evaluations could strengthen the case that stable targeted personality change is achievable.

Contextualizing Personality Judgment: Reading People in (and) Their Situations

FRANÇOIS S. DE KOCK

University of Cape Town, Cape Town, South Africa
francois.dekock@uct.ac.za

Abstract: Lievens presented a strong proposal for better integrating methods developed in the personnel selection field into personality research. However, a contextualized approach to personality assessment will have to understand better how people judge other persons, situations, and person–situation interaction to explain personality perception more fully. Dispositional and situational reasoning can be important pieces of the puzzle of how perceivers achieve accuracy in judging people and situations. I sketch a model of personality and situation judgment and highlight its practical application for moving the field forward. Copyright © 2017 European Association of Personality Psychology

Lievens presented a strong proposal for better integrating simulation methods developed in the personnel selection field into personality research. I agree that these methods may help to advance the important goal of contextualized measurement in personality assessment. By their very nature, situational judgment tests (SJTs) and assessment centre exercises (ACs) have as their *raison d'être* control and manipulation of situational characteristics to elicit particular applicant behaviours. This is a welcome challenge to the field, as personality and selection approaches share many aspects relevant to context-issues and may well be compatible.

Lievens's article, however, mainly focused on contextual assessment of personality. Lievens's proposals did not go far enough in contextual aspects relevant to personality perception. This is important as there is a trend towards reliance on personality judgments—rather than self-report personality

inventories—in both personality psychology research (McCrae & Weiss, 2007) and personnel selection (e.g. Zimmerman, Triana, & Barrick, 2010). My purpose here is to delve deeper into how dispositional and situation reasoning show potential for studying contextualized personality judgment. As we move in this direction, consider the following points, which are intended to enrich Lievens's argument in constructive ways.

The well-known 'personality triad' (Funder, 2006)—of persons, situations, and behaviour—has become mainstream thinking in what may *cause* behaviour, but it has not yet fully found its way into mainstream thinking about how we *judge* behaviour. In particular, the situations component (see Ziegler & Horstmann, 2015) requires better integration. Whereas the growing literature on dispositional reasoning has focused on declarative knowledge structures (or schemas; Fiske & Macrae, 2012) related to personality (Funder, 1999), it is not yet clear if similar knowledge structures may facilitate people's understanding of situations. Dispositional reasoning consists of three components, namely, induction, extrapolation, and contextualization. However, I see potential to incorporate perceivers' understanding of situations better into this framework. That is, these components can be extended to *situation reasoning* as explanatory variables for how we are able (or not) to judge situations correctly.

Drawing on a Brunswikian (1956) lens model framework, Figure 1 gives an example of how judgment objects (person traits and situation dimensions) are expressed in observable cues (behaviour and situation cues, respectively). As can be seen, to interpret these cues, perceivers must employ dispositional *and* situational reasoning as processing mechanisms with the objectives of making accurate trait inferences. For example, perceivers' understanding of situations is important to correct initial dispositional inferences in light of the situational context (Trope, 1986). To illustrate, a psychologist that correctly infers a competitive situation (dimension) from seeing candidates verbally 'jockeying for position' (cue) in a leaderless group exercise (situation) knows this behaviour is more likely to indicate assertiveness (trait) than (low) agreeableness.

Situational reasoning can be measured in *at least* five ways, informed by different literatures. First, extending Lievens's recommendation to consider situation strength (p. 433)—defined as the clarity and imperative nature of situational cues (Meyer, Dalal, & Hermida, 2010)—in AC design, I further propose that we develop measures to test perceivers' ability to identify situational strength (or intensity) correctly. Second, given that situation cues may be faint or unmistakably strong (on a continuum), cue sensitivity may be important. Signal detection theory approaches (Lord, 1985) have been used with success to measure cue sensitivity. Third, and somewhat related, perceivers implicit understanding of situational similarity (Klirs & Revelle, 1986) can shed light on perceivers' implicit situation theories, or their understanding of how situations go together and flow

from one another in real life. Fourth, emerging findings from personality research (Rauthmann & Sherman, 2017) show that perceivers are able to infer the dimensions underlying situations accurately. So we could build measures that present situational cues (to perceivers) and ask them to identify the most likely underlying situation dimension.[1] And finally, situation-trait relevance can be measured with the existing contextualization measure of De Kock, Lievens, and Born (2015). However, I agree with Lievens that this measure needs further refinement. Even though it shows useful measurement properties (De Kock, Lievens, & Born, 2017), the contextualization measure may be developed to distinguish between *levels* of relevance, rather than adopting dichotomous (relevant–not relevant) indicators associated with the multiple-choice question format.

The model I propose here has a few appealing advantages. First, it allows for more parsimonious and coherent explanations of how perceivers make sense of complex information emanating from persons, situations, and person–situation interactions. A theory of person and situation judgment that explains traits, situations, and how these interact provides richer hypotheses for future studies. Second, it proposes individual differences that facilitate person and situation judgment. I challenge others to develop measures of individual differences related to situation reasoning for use in further research. In particular, dispositional and situational reasoning measures show great potential for the practical study of contextualized personality judgment. In sum, my model may help deepen the understanding of how people make sense of others in the world.

Clearly, Lievens's proposals will stimulate and guide future research on contextualized assessment of personality. When faced with the challenges of understanding personality perception within situational contexts, researchers should consider whether methods developed in occupational and organizational psychology's selection research and practices may contribute to personality research. The personnel selection field—where simulation methods routinely manipulate situation characteristics—is a fertile testing ground for testing these ideas. Going beyond Lievens, I argued that
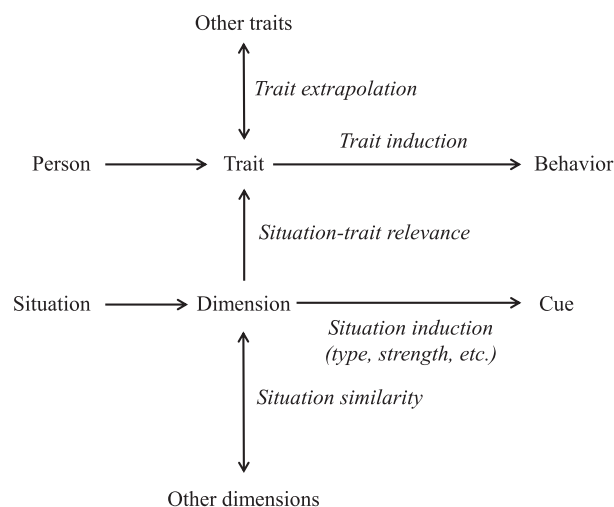

Figure 1.   A contextualized model of personality judgment in situations.

[1]However, my proposal rests on the assumption that 'objective' situations exist. Situation cues may also be diagnostic of multiple situation dimensions (see discussion of this in Rauthmann, Sherman, Nave, & Funder, 2015).

dispositional and situation reasoning measures can be used to produce rich and insightful research on contextualized personality perception. I believe such a view of person perception makes personality even more contextualized. Also,

practical ways to use these measures are suggested. I hope that my comment is helpful, and—beyond the critique raised above—I believe that Lievens's article offered an interesting perspective for future studies to consider.

Two Caveats to the Adoption of Personnel Selection Methods within Personality Psychology

WILLIAM L. DUNLOP and CALEN HORTON

University of California, Riverside, CA, USA
william.dunlop@ucr.edu

Abstract: Lievens has done the field a great service by underscoring the relevance that methods in personnel selection hold for personality psychology. We offer two caveats regarding his work. First, the methods Lievens outlined focus on personality traits. There exists, however, much to personality beyond traits, and these additional characteristics must not be ignored. Second, although the self-report measures of behaviour Lievens described require fewer resources than measures of actual behaviour, the added effort needed to study the latter is well warranted. Copyright © 2017 European Association of Personality Psychology

The second author of this commentary spent several years living in the Midwestern United States. This left him keenly familiar with the concept of silos. In his compelling article, Lievens hinted that he is also familiar with this concept. In this target article, he proposed that, largely unbeknownst to most in our field, those who specialize in personnel selection have been engaging in efforts that may be of benefit to our personality researchers. Perhaps, this is just the nature of academia, where knowledge is often 'siloed' away behind various disciplinary walls. In the particular case Lievens highlighted, however, there exists a substantial difference between what *is* and what *should* be.

We could continue to heap praise on Lievens and his thought-provoking article. Given that our task is to offer substantive commentary on his work, however, we have chosen to outline two, and only two, caveats we believe relevant to his work. First, although the methods Lievens outlined have traditionally focused on personality traits, these measures can and should be extended to include non-trait-based personality characteristics (e.g. motivations and narrative processes). Our second caveat is more of a cautionary note. Relative to assessment centre exercises (ACs), situational judgement tests (SJTs) require far fewer resources to implement. Given this disparity, we caution personality psychologists against shying away from adopting methods such as ACs, which are intended to assess actual behaviour within controlled settings (e.g. testing centres and university laboratories) in favour of self-report measures of behaviours.

PERSONALITY-SITUATION DYNAMICS: BEYOND TRAITS

We understand Lievens's main contention to be that personality psychologists have much to gain by adding to the mix methods common to the personnel selection field. In particular, he suggested that such adoption may hold

relevance to understanding 'within-person variability, trait-behaviour links, personality disorders, and personality expression and perception' (p. 424). We are certainly sympathetic to this position but believe that it would be strengthened by recognizing the fact that there is much to personality beyond the trait-based elements on which Lievens has chosen to focus. McAdams (1995), for example, contended that traits represent but one of three conceptual levels of personality, with the additional two levels encompassing motivational and contingent aspects of the person (Level 2: characteristic adaptations) and the integrative narrative of the self (Level 3: life stories).

Recently, the first author of this commentary (Dunlop, 2015) outlined benefits of adopting more contextualized approaches to the study of the personality characteristics indigenous to each of these conceptual levels. We will not rehash the entirety of this review here. Rather, we flag the fact that the SJTs, which Lievens has outlined, appear, at least to us, easily modifiable to allow study of personality dynamics at additional levels of personality. For example, when completing these tests, a respondent could be prompted to identify the motivations or goals he or she would have in the given hypothetical scenario (relevant to Level 2) or the framework (i.e. 'story') he or she would endorse to explain this scenario's significance (relevant to Level 3). In short, on the basis of everything we know about personality, in particular the fact that it is best represented by three conceptual levels (Dunlop, 2015; McAdams, 1995), we do not believe that the applicability of SJTs has been optimized.

SITUATIONAL JUDGMENTS VERSUS SITUATIONS THEMSELVES: A PITCH FOR STUDYING ACTUAL BEHAVIOUR

As might be evident, we see the inherent value in SJTs. That being said, one invariably makes huge leaps when drawing

inferences about *actual* behaviours from self-reports of behaviours. Indeed, many social psychologists have made careers by showing, time and time again, that there exist sizable gaps between the manner in which people think they will behave and the ways they actually do (for a humorous anecdote regarding the importance of studying actual behaviour, see Aronson, 2010). Thus, although SJTs possess a certain appeal in their economy, personality psychologists should not overlook the possible benefits that may be accrued from studying actual behaviours within controlled settings, as is done with ACs.

But what exactly might a researcher gain by forsaking quick and dirty self-report methods in favour of the more resource-heavy study of actual behaviours? To help answer this question, we offer the 'case study' of our colleague David Funder. Funder has dedicated much of his career to in-depth study of social interactions between participants across a number of settings or situations (e.g. a series of unstructured interactions between unacquainted persons; see Funder & Colvin, 1991). Although these data have taken considerable amounts of time to collect, code, and quantify, they have ultimately appeared in journals such as the *Journal of Personality and Social Psychology* (Funder & Colvin, 1991), the *Journal of Research in Personality* (Sauerberger & Funder, 2016), and even the *European Journal of Personality* (Morse, Sauerberger, Todd, & Funder, 2015). By investing in the more resource-intensive study of actual behaviour rather than the study of behavioural self-reports, researchers will likely be able to produce studies with greater cumulative impact.

WRAPPING UP

If the history of personality psychology has taught us anything, it is that the relation between persons and situations is complex. Because of this complexity, we need any and all possible empirical tools available to us. Lievens has done the field a great service by entering our silo and bringing a fresh perspective and novel assessment procedures that, if adopted, cannot help but extend our understanding of personality and personality dynamics. We encourage others to heed his call and follow his trailblazing efforts.

To Learn Something New, Try Something Different

C. EMILY DURBIN[1] and BRIAN M. HICKS[2]

[1]Michigan State University, East Lansing, MI, USA
[2]University of Michigan, Ann Arbor, MI, USA
cemilydurbin@gmail.com

Abstract: We commend Lievens for providing a compelling example of how diverse methods can enrich personality science and for highlighting the strategic utility of expanding our methodological reach. We describe some key advantages of this approach and note some novel ways these tools could be used to advance understanding of individual differences. Copyright © 2017 European Association of Personality Psychology

Lievens provided a thoughtful description of modern approaches to assessing individual differences in personality relevant to personnel selection and offered generative ideas about how to extend this work to advance other domains of personality science. These techniques rely on a very different frame for conceptualizing personality processes than the methodological model underlying broadband self-report assessment. Evidence of their predictive validity highlights the utility of this frame and its potential for illuminating how traits 'work' in ways that are obscure in traditional self-report approaches. Their potential lies in taking a narrow but deep approach that contrasts with the ubiquitous approach of assessing the broadest (and typically least contextualized) manifestations of traits. These approaches are narrow because they assess personality features embedded in and particular to important life contexts (i.e. work situations), and they are deep because they are designed to elicit rich evidence of trait processes as evidenced in behaviour and hypothetical behavioural choices in response to fairly delineated situational cues.

We suggest that traditional approaches of strip-mining personality via broad but relatively 'thin' strategies of assessment have reached a point of diminishing returns in generating new knowledge, and at this stage in our field, it is prudent to turn to other means that allow us to address new questions about personality. Extending the detailed and deeply contextualized methods Lievens described beyond the field of personnel selection is long overdue. Hopefully, his article will provide inspiration to personality scientists working on other aspects of individual differences and other life contexts in which personality matters to develop creative, ecologically valid, and theoretically informed approaches to studying mechanisms of personality.

Doing new things will require thought about to what degrees new approaches should yield familiar observations. For those whose views of psychometric and validity problems have been informed by the literature on self-report, Lievens's discussion of some disadvantages of self-report and the different problems one encounters with observational approaches should be informative. Interpretation of reliability and validity metrics is informed by expectations

about how the measurement approach works and about the nature of the constructs it measures. Approaches other than self-report are often depicted as suffering from substandard reliability and manifesting different patterns of associations with external correlates than has been typical, but this viewpoint operates from the assumption that all methods will operate the same way as self-report. Those working with more contextualized methods have a greater tolerance for variance unique to different approaches and relatedly a different appreciation for the tight connection between how personality is manifested and the situational and psychological conditions in which it is manifested.

Patterns of thought, behaviour, and motivation that interest us can be revealed differently when we use approaches that construct and manipulate contexts than when we smooth over those contexts. Lievens describes the typical correlations between observer ratings of applicant behavioural dimensions and self-reported traits as in the range of .20–.30. This effect size is very familiar to those of us working with observational measures of child traits, and it is instructive to note that in I/O psychology, this magnitude of association is seen as expected and not particularly underwhelming, while in developmental psychology, it has been a source of heartburn and some defensiveness. Lievens's article is an excellent model of how to make more explicit our expectations about methods and how to match them to the profile of methodological advantages each has and the conceptual promise each holds for understanding something interesting about personality. Lievens noted, 'any situation can elicit expressions of many different traits to varying degrees, and people can behave similarly though motivated by different traits' (p. 433). It is time to see this as a feature of observational and laboratory research, not a defect, and one which carries potential for providing ecologically veridical representations of the complexity of human behaviour.

Lievens offered some suggestions for how personnel selection approaches could be extended to other personality science domains. To add to this generative list, consider modifications of the Situational Judgment Test approach, typically used to ascertain participants' idea of the 'best' or 'preferred' strategy for responding to a scenario of personal relevance. Aside from asking about scenarios from other domains (e.g. relationships), one could also widen this strategy to learn more about potential mediators between internal psychological dynamics and trait-relevant behaviour, such as schemas about or expectations regarding the likely outcomes of behavioural strategies. Employing a 'think aloud' procedure as participants respond to these scenarios could elicit evidence of individual differences in mental models of standardized scenarios, memories of similar situations, or interpretations of the reasons why different strategies are effective. This would help fill in our understanding of how perceptions of situations help to bind patterns of behavioural regularities (i.e. traits). Similarly, the Implicit Trait Policy score approach could be extended to understand another scaling issue—connections between judgments and behaviours. In the domain of personality pathology, it is possible that those with more maladaptive personality structures have greater disconnects between what they view as effective behavioural responses to situations and the behaviours they actually display (more so than individuals without personality pathology). Think-aloud procedures could be used in conjunction with other measures of functioning to test whether this is in fact the case (or if those with more personality pathology engage in different interpretations of structured situational cues). Addition of observed behavioural data would help to quantify just how such psychological scaling works and its flexibility to individual differences. Persons whose behaviour is more weakly predicted by the objective contextual press of situations are especially informative for understanding how situations work to elicit behaviour because they reveal individual differences in how situations are experienced.

Even small steps towards trying something new, such as the ideas suggested by Lievens, could yield important new avenue for personality science, including knowledge to inform and reinvigorate traditional approaches.

A Behaviour-Based Selection Method to Assess Individual Differences in Personality and Identify the Exceptional

WILLIAM FLEESON and AYAT HAMZA

Wake Forest University, USA
fleesoww@wfu.edu

Abstract: We concur that personality psychology can benefit from assessment experience gained in the personnel selection field. We propose that combining a winnowing method with an assessment method, both from personnel selection, could create a cost-effective study method. The suggested method would be behaviour-based but relatively low-cost. Researchers would start with an undifferentiated sample of participants but via successive situation-based selections, steadily reduce that sample to smaller and more extreme groups. These extreme groups would allow investigating a dimension of interest at relatively high power and relatively low resources. Copyright © 2017 European Association of Personality Psychology

Lievens's proposal is excellent. Personality research can certainly benefit from adapting and adopting the behaviour-based personnel selection approaches used in the personnel selection field. Lievens masterfully created an appropriate balance by acknowledging the work in personality already completed along those lines while pointing out the much deeper experience of those in the I/O field. Adopting some of these methods will increase the degree to which personality psychology can be a truly behavioural science (Furr, 2009).

It is well known that personality psychologists extensively use global (broad) self-report measures. Such measures are usually valid and reliable, with deep conceptual work behind them. Additionally, they are extremely cost-effective and convenient. Over MTurk, a researcher could assess 300 subjects on a 30-item questionnaire for less than $100 and receive all the data in 1 day. In contrast, a behavioural study of 300 participants could easily cost over $5000 and take hundreds of hours to conduct.

Global self-reports nonetheless have drawbacks. They are susceptible to self-serving biases and blind spots (Vazire, 2010). Some may be based more on semantic memory than on specific episodes (Robinson & Clore, 2002). Because they are global, they typically do not illuminate how people actually behave in specific and real situations well. While global self-reports are rightfully held in high regard, there is an uneasiness about them that drives researchers to develop behaviour-based methods.

The advantages of behaviour-based assessments do not lead to universal adoption primarily because of their implementation cost and difficulty. Participants need to be paid $10 to $25 per hour. Trained personnel are needed to run sessions and code data. Recruiting, scheduling, running, and coding can require weeks or months. Assessment centre exercises involve even longer assessments of each participant, because of the recognition that multiple behaviours need to be assessed to get accurate assessments of individuals.

We propose a way to increase behavioural assessment in personality psychology while reducing many of these costs. This method builds on the assessment centre concept by appending another advance from the employment world, increasingly selective winnowing via successive cuts. That is, personality psychologists might adopt a selection method from the personnel selection field as well as an assessment method.

We call the proposed method 'Successive Situation-Based Selection'. In this method, the researcher has a dimension of interest, such as morality. The researcher determines a series of behavioural tests that tap this dimension, such as a series of ethically challenging situations like the dictator game or honesty tests. The method has a selection stage and an investigation stage. The researcher starts with a full and large sample of individuals (e.g. 200). The full sample completes one of the shorter behavioural tests. Only the participants who perform at the highest and lowest levels on the behavioural test are invited to the second session for another behavioural assessment. For example, in the dictator game, those who donate a lot of money or very little money would be invited to the second session. After the second session, only those who performed at the highest and lowest

levels in it are invited to the third session. This is repeated for a third behavioural session, which concludes the selection stage of this method.

The researcher will now have two small-sized groups. One extreme group who performed at the highest levels in all three sessions and another group who performed at the lowest levels in all three sessions. These two groups will participate in the investigation stage, in which the researcher investigates his or her empirical questions of interest. For example, the researcher can conduct experiments on the two groups. The researcher can administer questionnaires or interviews to learn about the two groups. The researcher can investigate the measurement properties (e.g. reliability and validity) of the behavioural tests used to select the individuals.

This method has several advantages. First, it assesses individuals behaviorally but for much less cost and inconvenience than assessing the full sample on everything. By successively reducing the sample size, researchers can save 50% or more of the cost and assessment hours. If situational judgment tests are incorporated into the selection phase, the cost would be reduced still further. A second advantage is that it increases the study's power per participant. Because the final participants are at the extremes, they contribute much more power to the experiment than would participants in the middle of the scale. A third advantage is that this method would allow researchers to focus on groups of interest. For example, in studies of morality, intelligence, achievement, need for cognition, and other similar concepts, this method helps to find the high-level individuals that are often of particular interest. This also applies to the low end in studies of concepts such as psychopathology, inhibition, and the 'dark triad'. A fourth advantage is that it builds on extensive expertise in the industrial-organizational field. A final advantage is that the method uses multiple behaviours to select individuals, avoiding problems associated with the unreliability of single behaviours for assessment.

This method also has disadvantages. Although winnowing reduces the cost, the method still costs noticeable amounts of money and time. Second, the method does not reveal the shapes or sizes of effects for the middle of distributions, such as whether effects are linear or non-linear. It reveals only the shapes and sizes of effects for the extreme groups. Further research will be needed to investigate the effect shapes and sizes in the middle of the distributions (when this method reveals an effect, it will thereby enhance the interest value of actually conducting a study of the middle of the distribution).

We are excited by the advantages of this method and see it as another way to build on the expertise in the industrial-organizational field to enhance personality research and address interesting theoretical questions.

ACKNOWLEDGEMENTS

Potentials and Limitations for Integrating Industrial/Organizational and Personality Psychology

DAVID C. FUNDER

University of California, Riverside, CA, USA
funder@ucr.edu

Abstract: I applaud Lievens's promotion of the use of advances in research on personnel selection by basic personality psychology. However, I evaluate his three key recommendations differently. First, I am enthusiastic about bringing assessment centre techniques into the toolbox of personality measurement. Second, self-report personality inventories might be enhanced by adding items aimed at measuring behaviours and attitudes, but this approach is not entirely new. Third, while making assessment instruments more context-specific may indeed improve predictive validity, excessive contextualization can undermine the explanatory value of personality trait constructs. Copyright © 2017 European Association of Personality Psychology

I have long been an admirer of Industrial/Organizational psychology's research in personnel selection,[2] as well as perplexed at its neglect by researchers in the highly relevant basic field of personality psychology. The most distinctive strengths of I/O psychology are (1) its focus on real-world outcomes and behavioural measures and (2) its quantitative sophistication. Whereas personality as well as social psychology have been for too long too fixated on the convenience of data gathered by questionnaire answers and keyboard clicks (Baumeister, Vohs & Funder, 2007), I/O research often includes dependent variables such as dollars of sales, days of missed work, longevity of employment, and supervisor satisfaction—all of which (and perhaps especially the last-named) are outcomes of interest to the supervisors who pay the bills. And it is I/O researchers who have made some of the most important contributions to methods such as factor analysis, meta-analysis, and psychometric assessment.

So I applaud Lievens's goal, which generously acknowledges the contributions basic research has made to the practice of I/O psychology and outlines an agenda for I/O to repay the debt. However, not all parts of the agenda are equally promising, and one part actually worries me a bit because even though the fields are relevant to each other, the ultimate goals of personality psychology and I/O psychology are not the same. In this comment, I focus on three aspects of the suggested integration of I/O with personality psychology and explain briefly why I love the first one, like the second, and am concerned about the third.

First, I am enthusiastic about the prospect for exploiting I/O psychology's assessment centre techniques and other clever means of behavioural observation. Within personality psychology, Peter Borkenau pioneered video recording of participants doing such things as singing a song, telling a joke, and reading newspaper headlines out loud (e.g. Borkenau, Riemann, Angleitner & Spinath, 2001). My research and that of others has observed participants doing diverse tasks such as building a toy, playing a game, or simply getting acquainted. It turns out that behaviour in such settings is closely and meaningfully related to self-reports of

personality, peer-reports of personality, and life outcomes (including health outcomes). But this kind of video-recorded behavioural observation has never really been systematized by personality psychologists. Systematic development of behavioural settings and measures that could, together, reveal aspects of personality missed by self-report scales is a task for which I/O psychologists have the relevant experience and expertise and therefore potentially much to contribute.

Second, Lievens explains how self-report items might be improved by lessening self-presentation concerns (or 'faking') by asking participants to describe behaviours (or 'behavioural preferences', which could be conceptualized as attitudes) rather than traits. But I would also observe that this approach is not exactly untried, even if it has not been systematized very well. For example, the latest version of the NEO-PI includes behavioural items such as 'When I'm around people, I worry that I'll make a fool of myself' and 'I sometimes get into arguments' and attitudinal items such as 'I believe variety is the spice of life' and 'I like loud music' (McCrae, Costa & Martin, 2005). These items do not ask participants to rate their own traits, though they are mixed in with items that do. Systematic investigation of the relative fake-ability and predictive validity of items in which people report their behaviours, their attitudes, and their traits would be a worthwhile but hardly radical innovation.

Third, Lievens points out how contextualizing assessment can improve predictive validity. It is indeed interesting to learn that items become more predictive of work performance when the self-report items include the phrase 'at work'. In more basic domains, it might be useful in a similar way to experiment with adding phrases such as 'at home', 'with my family', 'when under stress', and so forth. By the same token, it would not be surprising, and it could be useful, to demonstrate that behavioural observations made in a family context predict family-related outcomes better than behavioural observations made at work and vice versa.

But here is where the agendas of applied and basic science diverge. From an applied standpoint, one certainly would want to assess a predictive variable as precisely attuned to the outcome of interest as possible. But from a basic science standpoint, overly close attunement becomes nearly useless. This is because the goal of basic science is understanding,

---

[2]For brevity, I will refer to research in personnel selection as I/O psychology, though I/O psychology encompasses more than this topic.

not prediction. While the latter benefits from narrow contextualization, the former does not. If I may quote something I wrote elsewhere (and long ago):

> The use of narrow constructs may well increase correlations when predicting single behaviours, just as at the same time (and equivalently) it decreases the range of behaviours that can be predicted (Fishbein & Ajzen, 1974) … [but] explaining behaviour in terms of a narrow trait relevant to it and little else represents an extreme case of the circularity problem sometimes (unfairly) ascribed to trait psychology in general. If 'social skill at parties' is a trait detected by measuring social skill at parties and then is seen as a *predictor* or even *cause* of social skill at

parties, it is obvious that psychological understanding is not getting anywhere (Funder, 1991, p. 35).

Of course, I/O psychologist and employers aim to maximize the predictive validity of their measurements and everything else is secondary. Things are not so straightforward for a basic personality psychologist, who wants to understand why people behave as they do. Global traits, ones that transcend context, are useful explanatory tools precisely because they relate what people do in one time and place to what they do in another. Contextualize them too much, and they disappear.

Exploring the Interpersonal and Dynamic Nature of Persons and Situations through Assessment Centre Methods

ALYSSA M. GIBBONS[1] and DEBORAH E. RUPP[2]

[1]Colorado State University, Fort Collins, CO, USA
[2]Purdue University, West Lafayette, IN, USA
alyssa.gibbons@colostate.edu

Abstract: We agree that situational features are critical for understanding personality expression and that assessment centre exercises provide powerful tools for such explorations. Further, we argue that the interpersonal and dynamic aspects of these exercises provide unique opportunities to study how situations may be shaped by participants themselves, which is critical for understanding how personality is expressed in social contexts. Thus, we recommend that this design be expanded to assess momentary person–situation interplay, which would advance general understanding of such phenomena. Copyright © 2017 European Association of Personality Psychology

Assessment centre exercises (ACs) have considerable potential benefits to offer both selection and personality research. As Lievens pointed out, ACs allow participants to respond to explicit situational cues behaviorally across a wide array of contexts; allow exploration of dispositional reasoning; and offer potential to disentangle personality expression and perception. Here, we propose that ACs offer another key advantage worthy of scholarly investigation. That is, by their interpersonal nature, ACs are uniquely suited to further our understanding of the social and dynamic characteristics of both situations and personality.

Social features of situations are critical for understanding personality expression, as the same behaviour enacted in different situations may be interpreted as indicating different traits (Kammrath, Mendoza-Denton, & Mischel, 2005). For example, a person may be viewed as honest in speaking bluntly to friends, authoritative in speaking bluntly to subordinates, and rude in speaking bluntly to strangers. The inferences become more complex when the pattern is inconsistent; a person who speaks bluntly to friends but not to strangers is quite different from a person who speaks bluntly to strangers but not to friends (Gibbons & Rupp, 2009).

ACs often require social interaction (e.g. group discussions and role plays; Thornton, Rupp, & Hoffman, 2015). This creates opportunities to examine social characteristics of situations, such as numbers of people present, status of those persons, whether participants have cooperating or competing interests, and whether essential information is shared

or distributed (Thornton, Hanson, & Rupp, 2017). Although such features are not often systematically manipulated in ACs, some researchers have done so, with mixed results (Highhouse & Harris, 1993; Schneider & Schmitt, 1992). Given the advances described by Lievens, it seems an opportune time to revisit this line of research more systematically.

Such a line of inquiry would need to consider the dynamic nature of situations—that *features of the situation may be shaped by the participants themselves over the course of an exercise*. For example, a group discussion that was designed to be cooperative may become competitive if one participant perceives it as such. Similarly, a naturally arising conflict between two participants may create a new problem that must be resolved to complete an exercise successfully. This presents challenges for researchers using ACs to explore and manipulate situational cues, as these dynamic interpersonal factors may override other intended situational cues. This may explain why prior attempts to study the situational features of AC exercises have met with limited success: *because the participants themselves shape and even create the situation*.

At the same time, these spontaneous dynamics also present unique opportunities to understand *if … then …* patterns of behaviour in social contexts, as well as interactions among personalities. The example above highlights micro-level features of situations—participants' responses to others' behaviour—rather than the broader contextual features, such as the

types or general contents of the exercises (e.g. group vs individual simulation). Following the oft-noted bandwidth-fidelity tradeoff (Ones & Viswesvaran, 1996), we suspect that these micro-features may be more important cues for specific behaviours than broader features of situations as wholes. Such questions could be explored via AC methods, and such micro-features could be used to understand how personality is expressed in dynamic situations (e.g. a person who displays high Agreeableness until presented with conflict).

This type of research would require that AC assessors be able to evaluate and document both behaviour and the situations in which it occurs. In addition, researchers would need a meaningful way to aggregate and interpret the results of such detailed data. In our view, both requirements are challenging but achievable. Assessors are frequently trained to note the context in which behaviours occur (*International Task Force,* 2015). In our own experiences with ACs, we have observed assessor feedback providing such context (e.g. 'After your first idea was rejected, you did not offer any others'; 'When the role player contradicted your assessment of the situation, you used more aggressive language'). If these patterns are repeated, they become precisely the kinds of *if ... then ...* signatures we might use to draw inferences about personality and to make predictions about future behaviour or job performance.

To detect behavioural signatures, assessors would also require means to categorize the situational features they observe, similar to the way they presently categorize discrete behaviours into broader dimensions (Arthur, Woehr, & Maldegen, 2000). The recently developed situational taxonomies in the personality literature (CAPTION: Parrigon, Woo, Tay, & Wang, 2016; DIAMONDS: Rauthmann, Sherman, Nave, et al., 2015) could facilitate such in-the-moment situational assessments (if they are in fact assessed in the moment and not as overall exercise characteristics). That said, the CAPTION framework does not directly address the interpersonal characteristics of situations, and DIAMONDS does so only very broadly (it includes only 'Mating' and 'Sociality'—the degree to which a person perceives that positive social interactions are possible). Other relevant dimensions of social context, such as relative status of the individuals interacting, formal and informal sources of power, and access to resources, might prove useful as well. The *Atlas of Interpersonal Situations* (Kelley, et al., 2003) offers a potential starting point, but it seems that there is work to be done to integrate interpersonal cues into taxonomies of situational characteristics fully. ACs could prove valuable empirical tools, whereby assessors' observations of behaviours and situations could be used to examine covariation among situational cues (similar to the work of Parrigon et al., 2016 but based on observers' ratings). Comparing assessors' and participants' views on the characteristics of AC situations could also be a promising research direction.

Overall, we strongly concur with Lievens that AC methods have much to offer personality research. We propose that the social and dynamic nature of AC exercises should be viewed as an opportunity, not a challenge, as researchers continue to seek answers to these important questions.

Broadening the Scope: Situation-Specific Personality Assessment with Behaviour Description Interviews

ANNA LUCA HEIMANN and PIA V. INGOLD

University of Zurich, Zurich, Switzerland
a.heimann@psychologie.uzh.ch

Abstract: Lievens highlighted opportunities of employing situational judgment tests and assessment centre exercises for assessing personality–situation interplay. To broaden the range of possible approaches, we offer the patterned behaviour description interview as an additional selection instrument and outline why it might be particularly useful for studying the expression of personality in specific situations. In addition, we anticipate that diversifying methods for personality assessment will open up new research questions such as which methods are most suitable for studying which aspects of personality. Copyright © 2017 European Association of Personality Psychology

Lievens described how selection instruments such as situational judgments tests (SJTs) and assessment centre exercises (ACs) can be adapted to study the interplay between personality and situations. While his article provides a good foundation for integrating selection instruments into personality research, we urge researchers to consider a third promising selection instrument: the patterned behaviour description interview (Janz, 1982). Below, we elaborate on why this is a valuable method for situation-specific personality assessment.

Similar to SJTs and ACs, behaviour description interviews are popular selection instruments that can predict performance across different domains (Culbertson, Weyhrauch, & Huffcutt, 2017; Klehe & Latham, 2006). Within this interview format, target persons are asked about their behaviours in previously experienced situations. Thereby, interview questions can be 'designed to measure the specific job-related behaviors that are presumed to underlie a particular personality trait' (Levashina, Hartwell, Morgeson, & Campion, 2014, p. 265). In this case, each

personality trait is measured with several interview questions, and each interview question refers to a specific situation in which behaviours associated with the respective trait are expressed (for an example, see Van Iddekinge, Raymark, & Roth, 2005).

Behaviour description interviews may be particularly useful for assessing personality–situation interplay for several reasons. First, they have an open-ended response format. In contrast to traditional SJTs, the interview does not provide any response options so that interviewees must generate descriptive responses to the presented situations (i.e. interview questions) themselves. While some have questioned whether SJT items actually require specific situations (e.g. Krumm et al., 2015), behaviour description interviews cannot work without them. In fact, they may be especially effective at reflecting situational manifestations of personality, given that interviewees' responses are tailored to the presented situations.

Second, both interviewees and trained interviewers serve as information sources in the behaviour description interview. This is because interviewees provide self-descriptions of their behaviours (and eventually thoughts and feelings) in given situations, which are then evaluated by interviewers using anchored ratings scales. In contrast to ACs, the interview allows interviewers not only to learn about interviewees' behaviours but also to gather information regarding how they 'approach a variety of settings, as well as [...] their motivations for choosing certain behaviours' (Raymark & Van Iddekinge, 2013, p. 428). Thus, the interview may also capture aspects of personality that reflect cognitions and emotions (i.e., that are less visible when observing only behaviour from the outside, as typically done in ACs).

Third, each interview question refers to an actually experienced situation, and all interviewees are asked the same interview questions. Thus, behaviour description interviews provide high levels of contextualization (i.e. referring to concrete situations with actual tasks and characters), while also maintaining high levels of standardization in the way stimuli (i.e. interview questions) are presented. Consequently, behaviour description interviews combine advantages from both SJTs (i.e. high standardization) and ACs (i.e. high contextualization). In Table 1, we expand on Lievens's (2017) comparison of self-report personality inventories, SJTs, and ACs by summarizing features of the behaviour description interview.

Lievens (2017) noted that adapting selection instruments for assessing personality expands methodological diversity in personality assessment, which could help address some of the key questions in personality research. Specifically, his article outlines how SJTs may be useful to study trait-behaviour links and person–situation variability and how ACs may be useful to study trait expression and trait perception and their interplay. Extending the scope of employable methods, behaviour description interviews offer further intriguing opportunities to address key areas of personality research. Concerning trait-behaviour links, Lievens (2017) explained how implicit trait policies as assessed in SJTs may help trace situation-specific behaviours back to traits. We suggest that behavioural description interviews can be used to capture trait–behaviour links by explicitly asking

Table 1. Comparison of personality self-report inventories, SJTs, ACs, and behaviour description interviews based on Lievens (2017)

Building block (Lievens & Sackett, 2017)	Personality self-report inventories	SJTs	ACs	Patterned behaviour description interview
Stimulus format	Generic	Verbal descriptions of situations	Actual simulated situations	Verbal descriptions of situations' key characteristics
Stimulus presentation consistency	High levels of standardization	High levels of standardization	At best medium levels of standardization (predetermined cues are built in the exercises)	High levels of standardization
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.)	High levels of contextualization (referencing to tasks and characters in previously experienced situations)
Content targeted	Behavioural tendencies ('typical performance')	Procedural knowledge about effectiveness of traits in situations	Actual trait-related behaviour in simulated situations ('maximal performance')	Trait-related behaviour in actual (previously experienced) situations
Response format	Self-reports	Multiple-choice responses	Open-ended 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.	Trained interviewers evaluate the target person's responses and make ratings on dimensions.
Information source	Target person	Target person	Trained assessors	Target person and trained interviewers

interviewees why they behaved the way they did in specific situations. Since this is a more direct approach, it may add insights above the ones to be learned from SJTs. We also see the potential of behaviour description interviews to address further research questions such as within-person variability (given that interviews ask about behaviours in a variety of different situations) and trait expression and perception (given that the interview is a social situation in which interviewers evaluate expressions of personality).

We also would like to highlight that methodological diversity creates new research questions such as whether these various methods of personality assessment measure different aspects of personality and relatedly, to what extent these methods can be used interchangeably. First and foremost, this requires a conceptual foundation that guides future research on different methods of personality assessment. For example, while traditional self-report measures capture how individuals perceive themselves (i.e. identity), ACs may capture how an individuals' personality is perceived by others (i.e. reputation; see trait-identity-reputation model by McAbee & Connelly, 2016). On a related note, AC research showed that AC dimension ratings corresponded to highly observable traits (e.g. Extraversion) but hardly corresponded to less observable traits (e.g. Emotional Stability; Meriac, Hoffman, & Woehr, 2014). This is not surprising given that ACs, by definition, focus on assessment of clearly observable behaviours. Accordingly, we call for systematic research that elaborates on conceptual foundations to explore which methods can best assess different aspects of personality.

Adapting selection instruments to personality research, as Lievens suggested, appears to be a very promising opportunity to gain knowledge on the interplay between personality and situations. Its uptake relies on future personality and selection research (hopefully involving collaborations) to understand better what these methods capture and how to make best use of them in personality research across fields.

Multimethod, Contextualized Personality Assessment

CHRISTOPHER J. HOPWOOD and WIEBKE BLEIDORN

University of California, Davis, CA, USA
chopwoodmsu@gmail.com

Abstract: Lievens asserted that personality researchers should (a) use multiple methods, in contrast to traditional over-reliance on self-report and (b) move past highly general and context-free assessments to more careful consideration of the situations within which personality predictions are made. These points are with reference to personnel selection settings using the broader framework of Trait Activation Theory. Like most personality researchers, we agree with these general points but suggest expanding this argument to include more comprehensive and theoretically embedded multimethod approaches to personality assessment. Copyright © 2017 European Association of Personality Psychology

We agree with Lievens that personality research would benefit from using multimethod assessment tools that more carefully accommodate situational context. We further agree that it can be useful to borrow from neighbouring disciplines, such as organizational psychology, to identify, test, and synthesize new approaches such as the two specific methods highlighted by Lievens. Here, we touch upon a few issues related to multimethod assessment and contextualized personality measures, which lead us to expand Lievens's argument to a more general petition for a more comprehensive and theoretically embedded multimethod approach to personality assessment.

MULTIMETHOD ASSESSMENT

Lievens emphasized shortcomings of self-report and stressed the relevance of behavioural and contextualized personality measures. In doing so, he recited a familiar refrain: self-report is easy and cheap, but it is over-used and suffers from response bias and other method effects. Lievens's line of argument is a common narrative in the personality assessment literature: (a) self-report is a historical benchmark, which we used because it was easy and we lacked other tools; (b) however, self-report is overly limited and simple; and (c) the future requires new methods which provide hope for deeper understanding of what is really going on; therefore (d) we should validate new methods (using self-report as a criterion); and (e) show that the new methods provide incremental information over self-report.

This rationale de-emphasizes that self-report often ends up being more reliable, explains more variance in methodologically neutral outcomes, and is much less expensive than many new methods. The treatment of self-report in this story seems thus a little unfair. Rather than setting up competitions across methods, we see opportunity in the fact that methods can diverge. Low convergent validity can be informative, particularly if the factors that affect one method but not others can be identified (Bornstein, 2009), because this pattern suggests that various measures may be validly tapping different aspects of similar constructs. In other words, we do not think it is necessary to criticize self-report to promote other approaches.

This is particularly the case because it seemed to us that the situational judgment tests advocated by Lievens, while having some advantages relative to traditional trait

questionnaires, are self-report methods too. The assessment centre approach Lievens reviewed is not, but it is similar to other behavioural observation methods. While we generally agree that personality research could use more behavioural observation, we would also suggest that the reason it is not used more often is that systematic behavioural observation is challenging and costly to do, and its advantages may not always be worth the effort. Nevertheless, we fully endorse investing more time in developing novel observational methods and are hopeful that new technologies will make reliable and valid observational approaches increasingly accessible and useful.

In summary, we agree that a more diverse portfolio of assessment methods is needed to develop a comprehensive understanding of personality. However, despite our own enthusiasm for multimethod assessment (Bleidorn et al., 2017; Hopwood & Bornstein, 2014), we note that most of what we know about personality has come from self-reports. All measures are prone to error, and most methods are less reliable than self-report. Thus, we argue for more comprehensive multimethod approach to personality assessment that takes advantage of the benefits of multiple assessment tools including self-report.

CONTEXT

Personality psychology takes a holistic approach, and personality researchers have accordingly tended to focus on abstract constructs, without reference to situational or other contextual cues. As Lievens pointed out, adding more context to our questions and measures can enhance predictions considerably, if the predictions are made in that context. If we want to know how extraverted a person will be at a party, we should ask how extravertedly she tends to behave at parties in particular, rather than how extraverted she is in general.

We again join in Lievens's criticism but take a slightly different perspective. Personality researchers have been perhaps too focused on generalizability. For instance, it has become common to study whether 'personality changes' by administering a brief Big Five questionnaire multiple times and following its course. If no changes are observed, has personality been perfectly stable? We think a problem here is conflation of personality with a particular measure or set of constructs coupled with a relatively theoretical framing of the question. An alternative is to develop a more nuanced theory of what aspects of personality might change, and how those aspects might be most usefully organized and measured to answer specific questions (Wright & Hopwood, 2016). We offer a brief example from our own research.

Van Scheppingen et al. (2016) found that Big Five traits were not affected by the transition to parenthood, which was initially surprising. Further reflection made it clear that there are many possible explanations for this, including that only very specific aspects of personality might be impacted by this life event. For instance, recent longitudinal research observed that women experienced decreases in self-esteem and self-control following childbirth, likely due to challenges and stressors associated with transition to parenthood (Bleidorn et al., 2016; van Scheppingen et al., 2017). Broad Big Five measures might be too coarse to identify such patterns, highlighting the importance of measuring theoretically meaningful and contextually relevant traits.

In conclusion, we applaud Lievens's call for greater attention to context and promotion of multimethod assessment in personality research, find significant potential in the two specific approaches he advocated, and generally agree that personality researchers should borrow from organizational and other branches of psychology. But we add that the challenges of multimethod, contextualized assessment are familiar and chronic, and that progress will likely be incremental and slow. While situational judgment tests, assessment centre exercises, and Trait Activation Theory push personality research in useful directions, what is ultimately needed are comprehensive but flexible personality assessment models that can guide measure selection based on the specific issues being considered.

Could a Situational Judgment Test Predict Likelihood of Personality Change Following Experience of Adversity?

ERANDA JAYAWICKREME

Wake Forest University, Winston-Salem, NC, USA
jayawide@wfu.edu

Abstract: I discuss Lievens's arguments about applications of methods to assess trait expression in specific situations in personality psychology to the study of positive personality change in the wake of experience of adversity (post-traumatic growth). Specifically, adversity-relevant situations—and how individuals respond to such situations—may play important roles in whether post-traumatic growth ultimately occurs. Examining this question presents a test for the methods that Lievens advocated. Copyright © 2017 European Association of Personality Psychology

Lievens's excellent target article highlighted the valuable literature on personnel selection and performed a long-overdue service of identifying two approaches—situational judgment tests (SJTs) and assessment centre exercises (ACs)—that can stimulate new directions in personality research. I focus on one promising avenue: utilization of SJTs to predict likelihood of positive personality change following experience of adversity, post-traumatic growth (PTG; Jayawickreme & Blackie, 2014).

PTG AS POSITIVE PERSONALITY CHANGE

PTG has been defined as positive psychological changes resulting from struggle with highly challenging life circumstances (Tedeschi & Calhoun, 2004). It is conceptually distinct from resilience, which is generally defined as absence of negative outcomes (i.e. unchanged functioning) during or following potentially harmful circumstances (Infurna & Luthar, 2016). There is a robust literature examining retrospective self-perceptions of PTG, with individuals high in self-perceived PTG reporting greater appreciation of life, more intimate social relationships, heightened feelings of personal strength, greater engagement with spiritual questions, and recognition of new possibilities for their lives (Tedeschi & Calhoun, 2004). However, it remains unclear whether such PTG assessments correspond to actual changes in behaviour and cognition measured longitudinally (Blackie et al., in press) or improved levels of adjustment (Engelhard, Lommen, & Sijbrandij, 2015).

As I have noted elsewhere (Blackie & Jayawickreme, 2015), to the extent to which PTG is veridical (Fleeson, 2014; Frazier, Coyne, & Tennen, 2014), changes in situations that one experiences following adversity, as well as how responds to and selects specific situations may drive PTG. In a commentary on Jayawickreme and Blackie (2014), Jones, Brown, Serfass, & Sherman (2014) provided the example of how bereaved parents may experience changes in their daily situations that provide constant reminders of their recent loss of a child, which in turn may facilitate PTG. Whether such a parent experiences PTG, however, may depend on extent to which the parent has access to or selects specific situations in the wake of the loss of the child. For example, the parent may decide to seek out situations that offer social support, thereby strengthening improving quality of social relationships (Blackie & Jayawickreme, 2015). Of note, Mancini, Littleton, and Grills (2016) found that around 15% of a sample of 368 female survivors of the 2007 Virginia Tech campus tragedy experienced improvements in psychological functioning, and improvement among this sub-sample was moreover associated with increases in perceived social support and gains in interpersonal resources.

AN SJT THAT PREDICTS GROWTH?

Mancini et al. (2016)'s findings that people who experienced increases in psychological resources following the Virginia Tech shooting were those who also experienced increases in perceived social support and gains in social resources, as well as the importance of situation selection in the wake of adversity (Blackie & Jayawickreme, 2015; Jones et al., 2014) point to one possible question that methods Lievens highlighted could help answer. Specifically, an SJT could potentially identify individuals who would be more likely to experience positive changes in the wake of adversity (e.g. the group that experienced positive changes in Mancini et al., 2016). Such an SJT could not be easily constructed, since many adverse experiences are not made up of single situations but are more appropriately conceptualized as series of situations, such as unemployment (Lucas et al., 2003),

chronic illness (Tennen & Affleck, 1998), and daily hassles resulting from traumatic events (Miller & Rasmussen, 2014; see Blackie & Jayawickreme, 2015). Moreover, given that SJTs are based on assumptions that people encounter similar situations in daily life (as Lievens acknowledged), it is possible that a successful SJT would assess reactions to 'everyday' adverse life events, as opposed to unexpected trauma that is outside individuals' current life experiences. This presents a possible limitation of the SJT approach. Nevertheless, examining whether an SJT could identify likelihood of an individual drawing closer to others following experience of adversity presents a robust test of the method's utility.

Such an SJT could be validated against a prospective longitudinal experience sampling methodology (ESM) study (e.g. Blackie et al., in press) that additionally captures situations individuals experience in their daily lives. Participants would be administered the SJT at the beginning of such a study, and their responses would be validated against study results. The researchers would have to define a set of situations that would be relevant for such a study. To provide one example, while the DIAMONDS model (Rauthmann et al., 2014) includes adversity as a category, overcoming a highly stressful experience is likely to expand beyond feeling threatened (Blackie & Jayawickreme, 2015). One challenge here is that models such as DIAMONDS capture 'everyday situations', and many adverse life events do not fall into that criterion. Moreover, a distinctive feature of adversity is that it is imposed on individuals; individuals do not selecting. That said, individuals' responses to such adversity-relevant situations may determine whether PTG ultimately occurs; Mancini et al. (2016)'s findings provide support for the notion that individuals are not passive recipients of social support but instead actively seek to shape their social environments in accordance with their needs in the wake of adversity (Taylor, 2006; Mancini et al., 2016, p. 413).

A ROBUST TEST FOR THE SJT

Finally, developing an SJT for predicting who is more likely to experience positive changes in the wake of adversity represents a robust test for this method. I believe that the next step for researchers (such as Lievens) who want to adopt these methods for use in personality psychology should be to think critically about the types of questions that these approaches are best suited to answer. I hope that Lievens and others will present a list of specific critical questions in personality psychology that can be answered by employing these methods. Thinking through the utility of these methods for a question such as PTG presents one opportunity.

How Can Situational Judgment Tests and Assessment Centre Exercises Help Personality Psychology?

WENDY JOHNSON

University of Edinburgh, Edinburgh, UK
wendy.johnson@ed.ac.uk

Abstract: Lievens has offered personality psychologists some interesting suggestions for incorporating personnel selection methods into personality research. These can certainly spawn fruitful new contributions to specific aspects of the field. Ultimately, however, I think the most valuable contribution these methods might make might be to get personality psychologists to realize the limitations of traditional broad trait conceptualizations such as the Big Five. Copyright © 2017 European Association of Personality Psychology

Acknowledging their debt to personality psychology, Lievens has offered some interesting ideas for bringing into personality psychology methods and ideas personnel selection psychologists have developed for their purposes. They have found that assessing personality as well as their usual batteries of candidate experience and attainment resumes, reference letters, personal interviews, and so forth can enhance their ability to predict effective job performance. Personality psychologists maintain a veritable 'cottage industry' of studies reporting moderate associations between 'Big Five' personality traits and life 'outcomes' from body mass index and leisure-time activities to school grade-point averages and interpersonal relationship quality to life satisfaction and mortality, so mining the idea that what works for effective job performance will work for these other life outcomes too has merit. And Lievens has certainly made some valuable suggestions for doing this.

But, as he also notes, personality psychologists and personnel selection psychologists have very different core research agendas. Personality psychologists have as their primary goals understanding how our personalities are structured, how those structures emerge and evolve throughout life, and how whatever has emerged at any point is involved in our ongoing pathways. Personnel selection psychologists have as their primary goal helping people identify, among groups of self-selected candidates about whom only rather cursory and standardized information can generally be available, which would be best able to fill specific roles as they wish to see them filled. Moreover, before they apply the techniques Lievens discusses, at least some of the evaluators inevitably have sifted through a larger pile of preliminary written candidate information such as resumes, references, and stock application forms to select those on which to invest further evaluation, in the process likely forming sometimes-biased judgments that can affect appraisal of the subsequent evaluations (not to mention possibly eliminating excellent candidates). Personality psychologists face this potentially limiting step too but only indirectly.

I think this gulf between primary goals can tell both fields something important. Lievens's integration suggestions have focused on the aspects of personality psychology that involve situations most similar to the very constrained circumstances of personnel selection, and these are certainly important. But they also are rather 'fringe', special-interest questions in personality psychology. Attempts to integrate personnel selection methods into the core areas of personality psychology immediately bring the fraught questions of personality–situation transactions to the fore. The roles of the situation in personality and personnel selection psychology differ at least as fundamentally as do their overall research goals.

The selection setting is highly constrained at at least three levels. Each level is seen only rarely in real life and the combination almost never. One level is the activity script—what activities will be conducted, under what circumstances, and when. Another is evaluation of situation participants' specific behaviours within those activities. At neither level do the assessed participants have input, though they usually would. The third is more complex: participants' skin in the 'game' is huge, but it's skin in the overall assessment process—whether they'll be hired, which will be decided yea or nay on this basis alone and soon—and *not* in any particulars of the staged situations. That is, they have no ego investment, no reasons for any political or purely selfish motivations, backlogs of gratitude, resentment, or conditioned responses based on prior experiences in their specific assigned roles, nor need to live with the consequences of any staged situation itself: none of the 'baggage' that so often affects behaviour in day-to-day life. But they *are* highly motivated to display behaviours they *think* the prospective employer would want to see. The fact that personnel selectors know they need to worry about 'faking good' in their assessments indicates that often what contributes most to behavioural display is not the 'level' of personality 'trait' as defined in models such as the Big Five but the circumstances and choices that brought the person to the displaying situation and the person's interpretations of and goals within it.

This in turn suggests that personality traits as we define them are not latent biological 'forces', functions, or processes of some kind, 'caused' by particular combinations of genetic and environmental transactions that in turn 'cause' life outcomes the way rain causes grass to be wet or viruses cause fevers. Rather, they are behavioural patterns personality psychologists happen to have found very generally and arbitrarily useful to gather together to describe present population structure but little beyond this. These 'trait' collections each evolve idiosyncratically in each of us, directly in tandem with so-called 'life outcomes', in co-emergent rather than causal association. We don't exercise and keep our clothes neatly folded 'because' we are conscientious; rather we earn the label 'conscientious' when we accumulate lots

of orderly habits. Many of us exercise regularly but toss our clothes literally everywhere, while others are couch-potatoes with very neat dressers and closets. And lifetime regular exercise probably has a lot more to do with longevity than neat dressers and closets.

For personality psychologists, this means that efforts to improve 'life outcomes' by changing so-called personality 'traits' are unlikely to be fruitful. This also helps to explain the never-better-than-barely moderate correlations between those traits and job performance and other outcomes in both personality and personnel selection research. For personnel selection psychologists, this means they seek not people scoring highly on 'measures of extraversion or conscientiousness', but people who have the skills, motivations, and judgment to display appropriate behaviours falling under these banners in appropriate ways at appropriate times. And it limits the ability of personality-selection assessment exercises to help personality psychologists address their fundamental questions of how we come to be who we are at any point and where that leads us and to understand how situations are involved in this. Sure, use them where they can help us around the 'edges' of those questions, but if what we call personality 'traits' really do evolve as I suggest, what these exercises might accomplish most effectively is helping us see that the 'personality trait' emperor that has ruled the field for so long is wearing no clothes.

Situational Judgement Tests and Personality Measurement: Some Answers and More Questions

TIMOTHY A. JUDGE[1], JOERI HOFMANS[2] and BART WILLE[3]

[1]Ohio State University, Columbus, OH, USA
[2]Vrije Universiteit Brussel, Brussels, Belgium
[3]Universiteit Antwerpen, Antwerp, Belgium
judge.56@osu.edu

Abstract: Work psychologists have devoted considerable attention to studying how personality traits can best be conceptualized and assessed in 'high-stakes' contexts such as selection or hiring decisions. Lievens argued that two selection methods, Situational Judgement Tests and Assessment Centre exercises, by standardizing and contextualizing personality measurement, offer many advantages to personality psychology. In hopes of clarifying this argument, we ask two fundamental questions: (1) What aspects of personality do these methods fail to measure (are they deficient) and (2) What do they actually measure (are they contaminated)? Copyright © 2017 European Association of Personality Psychology

Filip Lievens made a general point with which we have full agreement, and several more specific points, which we argue, require further analysis and debate. Lievens's general argument was that personality and applied psychologists (who study selection decisions) each have much to contribute to the other's discipline. As researchers who have immersed themselves in both disciplines, we endorse this argument without qualification. Organizational psychologists have benefitted greatly from scientific progress in personality psychology, and, like Lievens, we believe personality psychology would benefit from a closer integration of research into how personality traits are studied in work contexts.

Lievens went on to argue that personality psychologists can learn from work psychology's research on Situational Judgement Tests (SJTs) and Assessment Centre exercises (ACs). He argued that these selection methods are worth considering because they (1) offer more precision and control than standard methods through standardization of the situations across participants and (2) have incremental validity above and beyond more traditional instruments. However, to appreciate their potential fully, we maintain that two important questions need to be answered regarding SJTs and ACs. First, what do we *fail to measure* by standardizing the situations? Second, what do *we actually measure* using these methods?

WHAT DO SJTS AND ACS FAIL TO MEASURE?

Lievens argued that SJTs and ACs, compared to experience-sampling studies, offer more control in testing within-person variability through situation standardization across individuals. Whereas we fully support the claim that research on within-person variability can benefit from more information on various sources of within-person variability, an important question is whether the best approach is to standardize the situations. In what follows, we argue that there are at least two concerns in doing this.

A first concern is that decades of research on situation selection has shown that people do not passively encounter situations but instead are actively involved in selecting, modifying, and creating them (Rauthmann, Sherman, Nave, et al., 2015). For example, Bolger and Schilling (1991) have shown that exposure to stressors accounted for one-third, and reactivity to stressors for two-thirds, of the meaningful variance in the relation between Neuroticism and distress in daily life. Moreover, through choice and enactment of situations conducive to one's personality, self-selection processes are believed to explain personality stability partially (Costa & McCrae, 1980). Because personality plays a key role in creating and enacting environments individuals experience, standardizing situations across individuals poses

several challenges. First, because people do not encounter situations randomly, standardization runs the risk of presenting people with situations that are not representative of those they typically encounter in everyday life. Second, it reduces personality to reactivity to (potentially unrepresentative) situations, which means that a significant part of the meaningful within-person variance is excluded.

WHAT DO SJTS AND ACS ACTUALLY MEASURE?

Whereas the previous question concerned construct deficiency (standardizing situations removes much meaningful variation in personality), one may also question the degrees to which SJTs and ACs are measuring what is actually intended. It has been argued that SJT items often have considerable heterogeneity at the item level, such that they show correlations with constructs that are not related to each other (McDaniel, List, & Kepes, 2016). Similarly, a recurring observation in the AC literature has been that scores on one dimension of an AC correlate highly with scores on other dimensions of the same exercise (i.e. low discriminant validity), whereas when people are rated on the same dimension in more than one exercise, there is little correlation among the dimensional ratings (i.e. low convergent validity). Although taking specific design recommendations into account can help boost their construct validity (e.g., Lievens, 2001), an ongoing concern is that the multidimensional nature of the presented situations in these assessments will always require assessees to respond on the basis of (the complex interaction of) multiple underlying traits.

Moreover, in addition to assessing multiple traits in the same context (i.e. a single AC/SJT exercise/item), we can also expect behavioural reactions evoked in these assessments to reflect constructs that do not belong to the personality domain. Indeed, one may argue that AC and SJT scores reflect procedural knowledge (job-relevant skills and knowledge) and cognitive ability as much as they do personality. Indeed, Lievens and Sackett (2012) argued SJTs can be '… viewed as measures of procedural knowledge in a specific domain (e.g., interpersonal skills)' (p. 460). Moreover, this confounding with non-personality constructs might actually explain why these measures are shown to offer incremental validity beyond trait measures (i.e. because they are, in part, not measuring personality at all). These concerns about contamination complicate usefulness of SJTs and ACs as 'pure' or unconfounded personality measures.

CONCLUSION

Like Lievens, we believe that personnel selection research has offered and continues to offer valuable insights into personality psychology. As our two questions suggest, however, we wonder whether the promise of SJTs and ACs as measures of personality traits is counterbalanced by significant concerns about their construct validity. Lievens welcomed further advancements on this front, and we believe one promising avenue is further investigations of the degrees to which SJTs both mediate and moderate relations between personality traits and work-related behaviour. Moreover, these methods might be valuable because they allow access to new constructs with which personality psychologists are less familiar. One example is implicit trait policies (ITPs), which tap into procedural knowledge regarding effectiveness of different trait levels. We believe that adding these constructs to the repertoire of personality researchers (or at least making personality psychologists think more about the criterion side) might help us to move a step closer to the ultimate goal: explaining why people behave, feel, and think the way they do. This is certainly a goal shared by personality and work psychologists.

Utilizing Advanced Psychometric Methods in Research on Trait Expression across Situations

JONAS W.B. LANG[1], JENNIFER L. TACKETT[2] and INGO ZETTLER[3]

[1]Ghent University, Ghent, Belgium
[2]Northwestern University, Evanston, IL, USA
[3]University of Copenhagen, Copenhagen, Denmark
jonas.lang@ugent.be

Abstract: Lievens emphasized the extent to which new measurement tools and additional statistics can be used to advance research on trait expression across situations. We suggest that advanced psychometric models represent additional important and complementary building blocks for progress and new insights in research on trait expression across situations. Here, we offer two specific examples of this: (1) Item Response Theory modelling of within-person variability simultaneous with estimation of latent trait levels and (2) estimation of latent trait and latent situation factors from a multitrait-multimethod framework. Copyright © 2017 European Association of Personality Psychology

We agree with Lievens that utilizing new measurement tools and gathering additional statistics such as within-person variability are important ways to advance research on trait expression across situations. However, we also believe that the field can be advanced further with even broader perspectives, including advancements in psychometric models.

Limitations of existing research can often be addressed by developing more advanced theoretical understandings of

existing measurement tools and psychometric models that formally test these advanced theories. Introduction of new measurement tools should ideally be closely aligned with development of psychometric models that capture the response processes in these new tools. We provide two examples of how psychometric models can potentially address common issues in research on trait expression across situations.

SITUATIONAL INCONSISTENCY

One challenging observation that persistently resurfaces in personality research is that people frequently do not consistently show the same trait-related behaviours across diverse situations. The behaviour thus appears to be inconsistent. In standard psychometric models such as the Rasch, 1PL or 2PL models, this type of variation is conceptualized as error variance. Accordingly, latent trait reliability commonly defined as the squared correlation between the estimated test scores from the psychometric model and the true underlying trait is low. However, more advanced psychometric models can dip into what standard psychometric models view as error variance, thus psychometrically advancing researchers' understanding of the role of situation. Observed consistency thus does not necessarily reflect psychometric consistency (Brown & Maydeu-Olivares, 2011; Lang, 2014). We frequently observe that researchers solely focus on observed consistency such as Cronbach's alpha or situation–situation correlations. However, what should ultimately be of interest are the true response processes that generated the data, and more complex models may be useful to uncover and measure these processes.

For instance, Lievens mentioned systematic individual differences in within-person variability as understudied sources of behaviour that may partially explain why behaviour is inconsistent. This suggestion could be a potential basis for psychometric explanations for inconsistency. In existing research, within-person variability is typically operationalized using within-person SD. Within-person SD is a potentially problematic measure when it is used for diverse situations because it (a) approaches zero when traits approach distribution maxima or minima, and (b) because it may capture not only systematic variation within-person but also other sources of variability such as individual differences in careless responding or individual differences in the standard error of measurement that varies across the latent trait continuum (from an IRT perspective). In most existing research on within-person variability, known limitations of the within-person SD are addressed by studying within-person variability over time in experience-sampling method studies using the same measurement scale so that within-person SD problems largely cancel out across time points.

One psychometric model that can accomodate situational inconsistency is an extension of IRT tree models and specifically Böckenholt's (2012) three-process model. This model separates Likert scale responses into three pseudo-items. One pseudo-item (the second in the tree) captures the trait direction and two pseudo-items (the first and third in the tree)

capture variability in people's responses to the items functionally independent from the trait direction (midpoint responding and extreme responding). Each pseudo-item is a separate ogive Rasch model defining response probability as a function of pseudo-item difficulty $\gamma_{\text{pseuditem } k, \text{ item } i}$ and a latent trait $\theta_{\text{pseuditem } k, \text{ person } j}$, $\Phi(\gamma_{k, i} + \theta_{k, j})$, where Φ is the cumulative normal distribution. Zettler, Lang, Hülsheger, and Hilbig (2016) used this model to study personality data, but it can readily be used for situational data using Likert scale-based assessments. By assuming one uniform variability trait for both variability pseudo-items, one might combine the two variability items. The resulting IRT model captures variability functionally independent from the latent (direction) trait and controls for item (situation) difficulty. This variability trait model is an attractive model because the variability trait has potential to explain additional variance in behaviour across situations, thereby reducing error variance, increasing reliability, and explaining why observed inconsistency occurs. However, the variability model is only one example, and the literature includes several other psychometric models that can uncover systematic traits from seemingly inconsistent behaviour such as, for instance, dynamic system models (Carver & Scheier, 1998; Lang, 2014; Revelle & Condon, 2015).

SITUATION FACTOR CONFOUNDS

A second observation that frequently surfaces in research on diverse situations and for which psychometric methods offer potential solutions is a pattern in which strong situation factors overshadow systematic trait variation. Research on behaviour in situations frequently utilizes the latent situation model, also known in the literature as the multitrait-multimethod (MTMM) model or more specifically the correlated trait, correlated situations model (Lance, Noble, & Scullen, 2002). The simplest version of this model requires assumption that response probability in particular situations are functions of situation 'difficulty' γ, the person's latent trait θ^T, and finally the person's latent tendency to respond in this particular situation θ^S, $\Phi(\gamma + \theta^T + \theta^S)$. This model can be estimated when it includes several traits and is typically used as a confirmatory test to estimate to what degree situations dominate over traits in MTMM data. However, the model can also be highly useful to model measurement of the latent situation factors as control variables so that underlying traits can be uncovered and estimated using factor scores. For instance, Tackett, Lang, Markon, and Herzhoff (under review) have recently used this model on 15 diverse situational ratings of the Big Five personality factors in children (also see Tackett, Herzhoff, Kushner, & Rule, 2016) and found that θ^T estimates had high latent trait reliability and that the pattern of intercorrelations changed after addition of θ^S controls. In other words, when systematic situational variation was added to the psychometric model, an underlying trait structure was uncovered. We conclude by noting that exploring advanced psychometric models to gather deeper understanding of trait expression across situations is likely one important building block for future research.

Promises and Pitfalls of the Personnel Selection Context in Disentangling Person-Situation-Behaviour Dynamics

SOINTU LEIKAS

University of Helsinki, Helsinki, Finland
sointu.leikas@helsinki.fi

Abstract: The potential usefulness of personnel selection contexts in personality/behavioural research is clearly and comprehensively argued by Lievens (2017). In my commentary, I (a) present some methodological concerns regarding this idea, (b) put forward how personnel selection contexts could be helpful in attempting to understand the roles of situations in personality/behavioural science, and (c) suggest how extending personnel selection research into work-place settings could benefit both personality and personnel selection research. Copyright © 2017 European Association of Personality Psychology

The experience sampling method (ESM) (e.g. Fleeson, 2001), and the empirical data on everyday behaviour that this approach has provided (e.g. Fleeson & Gallagher, 2009), is arguably among the most important developments in personality psychology in the last 20 years. ESM studies on personality have, among other important topics, brought the old (e.g. Frederiksen, 1972; Magnusson, 1981) but pressing (e.g. Reis, 2008; Wagerman & Funder, 2009) problem of the roles of situations into focus: where does the substantial amount of behavioural variance not encompassed by personality traits, or between-person differences, come from? How should we understand, categorize, and measure psychologically relevant situations or the features of these situations? How should we conceptualize and investigate the interplay between persons, situations, and behaviours? One thing seems clear: we need more data on actual behaviour, preferably from real-life situations.

In relation to these developments in personality psychology, Lievens (2017) presents a case for combining personality and personnel selection research, suggesting that selection contexts offer ways to examine within-person variability in behaviour and, thereby, the dynamics between personality, situation, and behaviour. This idea is appealing because, as Lievens (p. 425) points out, actual behaviour has a key role in such settings (although this seems to be mainly true of the assessment centre exercises—situational judgement tests only measure hypothetical behaviour, and while such measures can provide many insights, they should not be conflated with actual behaviour).

Despite its promise, combining personality and personnel selection research faces certain methodological challenges. As Lievens (2017, pp. 425–426) acknowledges, reliance on self-reports constitutes a limitation in personality research in personnel selection. It is well-established that personnel selection situations elicit socially desirable responding; that is, in such situations, people tend to present themselves in overly positive lights (e.g. Lönnqvist et al., 2007). This problem could be subverted by using observer-reports of personality; however, they may also be distorted because the application context itself is likely to elicit socially desirable behaviour. Furthermore, given that self-report is the method of choice in the bulk of personality research, using observer-reports obtained in a personnel selection situation would likely not allow for direct comparison with other personality research.

In sum, personnel selection contexts may not be optimal for conducting 'personality traits-predicting-outcomes' studies, unless personality ratings can be obtained from a source outside of the personnel selection context.

Lievens suggests developing contextualized measures of personality (p. 426) as a potential solution to the social desirability problem. Some issues should be considered before moving on with this suggestion. Although such variables may be valuable in personnel selection, they may, due to lack of theoretical and empirical underpinnings, be less useful from the perspective of basic research, especially if they are developed 'on-demand' for a particular study/selection setting. The Big Five/Five Factor model of personality is far from perfect, and the idea that broad, decontextualized traits are all we need in personality psychology can and should be called into question. However, (a) the Big Five/Five Factor model has an impressive amount of empirical support and (b) when we employ these traits, everyone more or less understands and agrees on what we are measuring. This may not be the case with contextualized personality variables, which often lack firm theoretical or empirical underpinnings. To justify contextualized personality variables, we need more data on situations, behaviours, and situation–behaviour contingencies.

Despite these concerns, I wholeheartedly agree with Lievens that personnel selection contexts could help gather valuable information on person-situation-behaviour dynamics, especially with regard to the roles of situations. Rozin (2009) famously noted that as compared to the natural sciences, psychology lacks basic behavioural data. Although there has been some improvement during the last decade, Rozin's statement is still very much true, and this omission may at least in part be responsible for the situation problem.

One way of tackling the problem of situations could be to obtain behavioural data from a broad range of everyday life situations and, based on such data, to formulate behavioural 'regularities' related to particular situations or features of situations; that is, formulating situation–behaviour contingencies (as has been done, for instance, in research conducted within the framework provided by interpersonal theory; Fournier, Moskowitz, & Zuroff, 2008). Assessment centre exercises could be especially useful in examining behavioural differences between such psychologically relevant situation pairs as competitive versus cooperative, dyadic versus group, or social versus non-social situations.

Finally, and in relation to the above points, the personnel selection situation is very particular in terms of rarity, pressure, and self-presentational concerns. Thus, while there is much to gain from studying selection situations, I would suggest moving beyond them and into the workplaces themselves. For instance, longitudinal research following employees from the personnel selection situation into their workplaces could measure employees' workplace behaviours as well as their occupational and social competence in different situations and obtain ratings or behaviour, skills, or personality from several sources (e.g. bosses, peers, and clients). Such research would be highly valuable for both personality and personnel researchers (and employers!). In addition, personality ratings could, in such research, be collected also after the selection situation, circumventing the problem of socially desirable responding.

In conclusion, as Lievens compellingly argues, personnel selection settings offer valuable contexts for studying situation–behaviour contingencies, but conducting basic personality research in these settings may not be the best course of action. Furthermore, for the mutual benefit of personality and personnel psychology, it would be desirable to extend behavioural and personality research from selection contexts to the workplace.

Applying Personnel Selection Techniques to the Psychological Study of Accurate Personality Judgment

TERA D. LETZRING and DOUGLAS E. COLMAN

Idaho State University, Pocatello, ID, USA
letztera@isu.edu

Abstract: We agree with Lievens's proposal to assimilate methods from personnel selection into the study of personality psychology and in particular into research on person perception accuracy. Researchers can apply Situational Judgment Tests to study judgment accuracy by using them as the criteria and judgment measures. Assessment Centre exercises could also be fruitfully applied to increase the fidelity/external validity of the accuracy criterion. Lastly, we touch on some key advantages and disadvantages of transferring personnel selection techniques to research on accurate personality judgment. Copyright © 2017 European Association of Personality Psychology

Lievens encouraged cross-pollination of methodological techniques between distinct, yet relatable, research literatures. He suggested that two approaches to personnel selection—Situational Judgment Tests (SJTs) and Assessment Centre exercises (ACs)—are ripe for transfer. We agree with this assessment and explicate ways that personality psychologists could use SJTs and ACs, specifically in research on accuracy of personality judgments.

Personality psychologists assess judgment accuracy as the agreement between a target person's actual standing on some personality characteristic(s) and a judge's assessment of the same characteristic(s). Targets' actual standing on a characteristic is usually assessed with self-report or a composite of self-reports and ratings from targets' acquaintances. Typically, assessments of personality judgment accuracy have focused on broad characteristics, without explicit consideration of how situations could influence these judgments. Therefore, we outline ways to apply SJTs and ACs to study personality judgment accuracy.

APPLYING SJT AND AC APPROACHES

Situation-specific SJT items could be used to assess accuracy. Targets would choose the behaviours from a list that they think they would perform in a set of situations, and responses would be the accuracy criteria. Ratings could also be gathered from acquaintances regarding what they think the targets would do, and items with convergence between self- and acquaintance-ratings could be used in the judgment task. After some exposure to the targets, judges would report what they think the target would do in the same situations. A judgment would be deemed accurate if the judges selected the same behavioural response as the targets and the acquaintances. This approach could be problematic because the accuracy criteria would consist of what people *think* they would do, and people are not always able to predict their behaviour accurately, particularly in novel situations. Therefore, it must be expected that the link between intention and behaviour would be relatively weak because assessment of accuracy would be based on how others *think* they would behave instead of how they actually behave.

External validity/fidelity of the accuracy criterion could be increased by using ACs to capture actual behaviours, and these behaviours would be the accuracy critieria for SJT items. The judges' task would remain the same as in the purely-SJT approach. The difference would be that response accuracy would be based on targets' actual behaviours.

In both approaches, judges select a single behaviour they think would be performed in a specific situation by a given target. As such, each item would be scored as correct or incorrect. A benefit of this assessment scheme is a straightforward computation and interpretation of accuracy—the number of correctly identified behaviours. This method may result in less variable accuracy scores than traditional measurement methods that assess accuracy as the similarity in ratings on Likert-type scales with correlation or regression approaches. This is important and potentially problematic because detection of true variation is essential to examining how accuracy relates to other constructs.

ADVANTAGES AND DISADVANTAGES

There are several advantages to applying the methods from personnel selection to research on accurate personality judgment. First, there would be greater methodological variance in accuracy research, which is beneficial for examining the generalizability of findings. Second, there would be a greater focus on the contexts in which judgments are made. Personnel selection focuses on job-related characteristics within employment situations, so situations are designed to elicit certain characteristics.

Person perception research could operationalize a similar intention by carefully thinking about the characteristics being assessed, whether behaviour is expected to vary across situations and which situations would best elicit relevant cues. For example, for judgments of extraversion, situations should allow for variability in amounts of talkativeness and assertiveness. Even for traits traditionally associated with lower accuracy, it is possible to create situations that elicit variability in relevant behaviours (i.e. anxiety-provoking situations and neuroticism; Hirschmüller, Egloff, Schmukle, Nestler, & Back, 2015).

Our discussion above speaks to the importance of considering more than one moderator of accuracy at a time (e.g. judge and information), which according to Lievens is consistent with recent assessment centre research. Researchers should also attend to the relation between good judges and good targets. To capture true accuracy variability among judges, good targets are needed because they make many relevant cues available and this is necessary for assessing

differences in judgment ability (Rogers, 2015; Rogers & Biesanz, 2016). An AC approach could increase the likelihood that targets would be 'good' by engaging in personality-relevant behaviours, thereby making it possible to detect differences in judgmental ability.

A third advantage comes from having an explicit purpose for the judgments. Judges in person perception accuracy research are typically not given a reason or motivation for being accurate, whereas in real-world situations, there is often a known reason for making judgments and motivation to be accurate. Providing research participants with rationale and motivation to be accurate may change the way they approach the task and therefore affect accuracy and how it relates to other variables of interest.

There are also disadvantages that should be considered. First, using ACs to create accuracy criteria would require much time and resources, as each target would be put in multiple situations on which to base judgements. This would require a large investment of time, and situations would have to be realistically created by using additional participants or research confederates. A possible solution to this disadvantage would be to create a standardized AC test. Creating such a test would require a large investment of time and resources, but this would yield a test with high external validity that could be used by multiple researchers.

In conclusion, we see merit in Lievens's proposal that personnel selection techniques could be fruitfully applied to personality research and specifically to research on personality judgment accuracy.

A Call for Cross-Fertilization between Personality and Personnel Selection Researchers

DANIEL V. LEZOTTE[1], DAVID M. CONDON[2] and DANIEL K. MROCZEK[3]

[1]Mercer U.S., USA
[2]Northwestern University, Evanston, IL, USA
daniel.mroczek@northwestern.edu

Abstract: Lievens made a case for using SJTs in personnel selection, a recommendation with which we agree. In particular, we like the emphasis on branching out from current methodologies and using new techniques such as SJTs not only in I/O or personnel selection research but also in basic personality research. Despite our enthusiasm, we point out several problems, most notably absence of time dimension in SJTs. Copyright © 2017 European Association of Personality Psychology

Lievens made an important contribution by laying out the case for more cross-disciplinary research collaboration between personality and personnel selection researchers. For any research discipline to make significant strides, it must branch out from methodologies and findings within its own discipline to learn from and share with other disciplines. This also applies to sub-disciplines of personality psychology and industrial/organizational psychology, both of which can learn from each other through cross-fertilization of ideas and methods. Lievens made a compelling case that this will likely extend utility of findings in both areas.

Selection researchers and practitioners rely heavily on personality trait theories and established methods for assessing these traits. However, as Lievens pointed out, in personnel selection, cognitive ability tests have consistently shown higher levels of validity (prediction of job performance) than personality tests. Meta-analyses on Situational Judgment Test (SJT) validity have shown them to predict less well in selection contexts than cognitive ability tests but better than personality tests (Schmidt, et al., 2016). Assessment centre exercises (ACs) are similarly situated between cognitive ability and personality tests, although their validity is generally higher

than that of SJTs. This is likely a result of ACs incorporating multiple methods, for example in-baskets, role-plays, leaderless group discussions, and more traditional written assessments. At best, personality tests can be said to have modest validities, probably due to breadth (facets, sub-traits, or even specific items may have higher validities than broad macro traits like the Big 5; Mõttus, 2016). This suggests there is considerable need for additional research on ways that personality traits can be more effectively deployed in personnel selection. A focal point for this research should be development of more predictive personality measures. Lievens suggested that SJTs are viable options for such development. Though SJTs are considered distinct from personality tests in practice, this is primarily because they do not usually measure single familiar personality traits (e.g. Conscientiousness). Instead, they often tap multi-dimensional constructs such as interpersonal skills; the highest validities have been found with SJT measures of teamwork and leadership (Christian, et al., 2010). However, it seems clear that personality traits underlie these broader measures.

In assessing personality, selection practitioners have been searching for alternative to self-report questionnaires for many of the reasons Lievens cuted, that is susceptibility to faking, applicant perceptions, and lack of contextual specificity. This has resulted in greater focus on video-based simulations that present realistic job scenarios, as well as gamification and game-based assessments, which are akin to generic, non-job related simulations. While Lievens suggested SJTs may be one alternative, they would need to be tailored for this purpose, as the majority of SJT's used for selection tend to assess broader characteristics than would be useful for personality research. However, it does seem plausible that personality researchers could benefit from developing such SJTs, and their implicit assumption that personality expression is fundamentally situational.

Lievens also argued that ACs could benefit personality research. Certainly, personality research could likewise benefit from AC development and implementation. He pointed out that personality researchers can tailor the AC approach to their own needs (e.g. school, health, relationships, and sports) and identified several intersections of research that could be explored by both selection and personality researchers. ACs are unique in that they use rater observations to assess personality traits of others, based on actual behaviours rather than self-reports or reports of actions 'one would

take' under different scenarios. Further research in both directions would benefit current AC trends, and personnel selection in general, and would be timely as many virtual (computer-based) ACs are being developed to capture the benefits of live ACs while making the process shorter, less time- and resource-intensive, and less expensive.

Though Lievens made solid and interesting arguments for using SJTs in personnel selection, some extensions of his points seem warranted. Presenting candidates with relevant job-related situational descriptions and assessing reactions to them can potentially illuminate competencies and capabilities that are crucial for high performance. Yet this approach is not without flaws. A critical aspect of the SJT technique is to obtain senses of ranges of situation- and trait-relevant behaviours expressed by individuals. Developmental quantitative psychologists refer to this as intra-individual variability (Nesselroade & Boker, 1994). It is extent to which someone expresses wide ranges of behaviour over periods of time. For example, some extraverts display extraverted behaviour over 90% of the time, while other extraverts display lesser amounts, say 60% to 70%. Of course, this can be construed as the former having higher absolute level of extraversion, but it also implies more intra-individual variability in the latter. Lievens acknowledged this but left out a critical component: time. In a contrived SJT that takes place over an hour or maybe a half day at the most, the temporal dimension is lost. Yet behavioural variability unfolds over time, over periods of days and weeks, even years. Characteristic situational reactions, too, unfold over such periods. Of course, it is difficult to carry out a personnel selection assessment over such periods! A practical barrier exists. However, the concept of time is critical and could be studied more in-depth among current employees to allow some predictive linkage between pre-hire assessments and post-hire time-based variability in behaviour and performance.

Another issue is one alluded to above when we used the word 'contrived'. SJTs, like social psychology lab experiments, may not bear much resemblance to real-life situations. Lab situations are often wanting in ecological validity. Many phenomena occurring in labs do not generalize to the real world, and this constrains their practical utility. SJTs may be enhanced by using situations that are as close to real life as possible. However, these criticisms are minor, and we generally like the approach Lievens articulated.

Breaking Down Institutional Silos: Forging Fruitful Linkages Between Industrial/Organizational and Personality Psychology

SCOTT O. LILIENFELD[1,2]

[1]Emory University, Atlanta, GA, USA
[2]University of Melbourne, Melbourne, Australia
slilien@emory.edu

Abstract: Lievens contended persuasively that personnel selection methods drawn from industrial-organizational literature have much to offer to personality and clinical psychology. I elaborate on his arguments by (a) conceptualizing situational judgment tests (SJTs) and assessment centre exercises (ACs) within an Allportian perspective on

personality, (b) discussing limitations of SJTs and ACs in imposing constraints on individuals' propensities to select, shape, and create situations, and (c) outlining how ACs may further inform assessment and understanding of personality disorders. Copyright © 2017 European Association of Personality Psychology

Lievens made a persuasive case that personnel selection methods can inform a variety of debates in personality psychology. I agree heartily with his thought-provoking arguments. For reasons that are partly institutional, personality and clinical psychology have remained largely insulated from industrial/organizational (I-O) psychology. I-O psychology is underrepresented in many major psychology departments. Even in those departments in which it features prominently, it is rarely well-integrated with other subdisciplines. Some of the schism may be conceptual too. Beginning with Spearman (1930/1961) and later Cronbach (1957), psychologists have bemoaned separation between the two great disciplines of scientific psychology, experimental and correlational. Although much of the I/O literature, including self-report selection methods, embraces the correlational tradition, some of the paradigms advocated by Lievens, especially assessment centre exercises (ACs), appear to fall squarely within the experimental tradition, placing applicants in manipulated situations that simulate those encountered in workplace settings.

It may be helpful to place situational judgment tests (SJTs) and ACs within the context of an Allportian perspective on personality (see Funder, 1991). Although complementary, these two paradigms capitalize on differing characteristics of personality traits. SJTs seem to fit comfortably within Allport's (1961) conceptualization of traits as dispositions 'having the capacity to render many stimuli functionally equivalent' (p. 347). Thus, via Piagetian assimilation, traits shape the way people perceive reality, organizing disparate stimuli into functionally similar equivalence classes (Bowers, 1973). The extraverted individual, for example interprets a host of situations, such an as intimate seminar, a club outing, or a trip to a new restaurant, as opportunities for socialization.

In contrast, ACs capitalize on propensity of traits to interact statistically with situational triggers, such as role-plays or presentations. ACs may thus elicit dispositions that otherwise remain dormant. Lievens wisely avoided the error of earlier scholars (e.g. Bowers, 1973) who presented this interactionist position as at odds with a trait position. As Tellegen (1981; see also Ben-Porath & Tellegen, 1990) observed, we can conceptualize most or all traits as inherently interactional, as they are 'if stimulus, then response' dispositions. The fact that narcissistic traits display higher response penetration (Tellegen, 1991) into behaviour following certain provocative situations, such as an insult from a supervisee, than other situations, such as an amicable meet-and-greet with co-workers, does not challenge the existence of trait narcissism. To the contrary, this statistical interaction is mandated by this trait concept.

One limitation of SJTs and ACs, briefly acknowledged by Lievens, is that by presenting participants with standardized situations, they short-circuit 'personality-driven situation experience' (Rauthmann, Sherman, Nave, et al., 2015, p. 96), the propensity of traits to influence which situations individuals encounter. This point may warrant elaboration.

One crucial way in which traits manifest themselves in everyday life is via the propensity of individuals to select and create situations (Ickes, Snyder, & Garcia, 1997). Rauthmann, Sherman, Nave, et al., (2015) distinguished several variations of this tendency, including (a) selecting situations, (b) shaping extant situations, and (c) creating novel situations. The high sensation-seeker, for instance, may (a) elect to join a sky-diving club, (b) persuade the club to attempt risker jumps, (c) establish his or her own skydiving club, or all three. In the midst of the person-situation debate, Wachtel (1973) argued that some of the modest cross-situational consistency of putative traits underscored by Mischel (1968) stemmed from researchers' insistence on a model of the 'implacable experimenter', in which researchers' actions in response to participants' behaviour is rigidly standardized. Similarly, much of the heritability of personality may stem from active genotype–environment correlation, the propensity of individuals with certain genotypes to select, shape, and create situations that suit their genetically influenced dispositions (Knafo & Jaffee, 2013; Scarr & McCartney, 1983; see also Bouchard, 2016 and Johnson, 2010 for discussions of Experience-Producing Drive Theory).

By constraining the situations to which participants are exposed, ACs may preclude full trait expression. In this respect, ACs may benefit from affording evaluators greater flexibility in allowing them to react spontaneously to participants' behaviour during interpersonal exercises. The loss in internal validity should be offset by gain in external validity generated by trait expression.

Finally, Lievens observed that SJTs may enhance detection of personality disorders (PDs), in part by eliciting relevant individual differences in interpretation of ambiguous scenarios. I concur and add that ACs may play useful roles here as well, as they may elicit distinctive interpersonal manifestations of many PDs. Beginning with early interpersonal theorists (e.g. Leary & Coffey, 1955), psychologists have recognized that many and perhaps all PDs are inherently interpersonal conditions (Wilson, Stroud, & Durbin, 2017). The concept of a narcissistic, paranoid, or avoidant PD would seem to possess scant meaning for describing individuals stranded on a desert island.

Implicit in interpersonal models is that PDs are malignant configurations (read: statistical interactions) of two or more traits (Grove & Tellegen, 1991). For example, psychopathic personality may reflect paradoxical co-occurrence of boldness (e.g. social poise and charm) on one hand and maladaptive traits, such as coldness (e.g. callousness and guiltlessness), on the other (Lilienfeld, 2013; Patrick, in press). This conjunction of typically contradictory traits can deceive observers, who may assume that psychopathic individuals are more benign than they are. ACs may offer helpful vehicles for testing such

interactional hypotheses, which posit aversive interpersonal outcomes only in the presence of specific Trait × Trait interactions.

One challenge with using ACs and related personnel selection procedures in this context is the dimension of time. Individuals with psychopathic and narcissistic PDs in particular often create positive first impressions that fade over time. For example, people with pronounced narcissistic traits tend to be better liked than other people on short-term acquaintance but more disliked on long-term acquaintance (Back, Schmukle, & Egloff, 2010). Psychologists may wish to harness this temporal discrepancy to their advantage by treating it as signal rather than as error. That is, a propensity to generate a highly positive impression in an AC that is not matched by effective performance later may be a marker of certain PDs. This adaptation illustrates merely one way among many in which personnel selection procedures can be harnessed to provide novel indicators of personality pathology.

Hiring the Good Target: Towards More Integration of Expressive Accuracy into Personnel Selection

MARIE-CATHERINE MIGNAULT and LAUREN J. HUMAN

Psychology Department, McGill University, Canada
lauren.human@mcgill.ca

Abstract: Lievens described methods and findings from personnel selection research that may benefit personality perception research. However, Lievens's suggestions primarily focus on assessors' ability to perceive candidates accurately, just as personality perception research has historically focused on the judge, overlooking the important role of the target in the person perception process. We therefore consider how research on the target, such as factors that promote expressive accuracy, could be applied to personnel selection contexts, which would in turn advance research in both fields. Copyright © 2017 European Association of Personality Psychology

Accurately perceiving candidates in personnel selection is likely essential to building high-performance, cohesive teams. Similarly, accurately perceiving potential friends may be essential to developing harmonious friendships. Can research on perceptions of potential employees benefit research on perceptions of potential friends? Lievens presented insights from personnel selection research that may indeed benefit personality perception research. However, one key element of personality perception yet to be fully integrated into personnel selection is the 'good target'—individuals who tend to be accurately perceived (Colvin, 1993a; Funder, 1995). We consider how work on good targets may enrich personnel selection research, while highlighting aspects of personnel selection that may inform work on the good target.

BRINGING THE TARGET INTO PERSONNEL SELECTION RESEARCH

As with much personality perception research, personnel selection research appears to have primarily examined accuracy through the assessor's lens. However, considering the roles of the targets of the assessors' judgments (i.e. candidates) may offer more room for improving accuracy. As Lievens noted (p. 435), recent findings in personality perception research suggest that there is more variability in how accurately people tend to be perceived, or expressive accuracy, than in how accurately people tend to judge others, or perceptive accuracy (Human & Biesanz, 2013). This may also be true in selection contexts, so selection research would likely benefit from considering factors that influence candidates' expressive accuracy.

ENHANCING CANDIDATE EXPRESSIVE ACCURACY

Contrary to Lievens's contention (p. 435), self-presentation or impression management does not necessarily obstruct accurate assessment. Rather, self-presentation often involves potentially conflicting motivations to be seen both positively and authentically (Baumeister, 1982; Leary, 1995; Schlenker & Pontari, 2000). Consistent with this, observers' tend to achieve accuracy even when judgments are based on stimuli from high self-presentation contexts, such as online social media profiles (Back et al., 2010). Self-presentation can even increase how accurately individuals are viewed (Human, Biesanz, Parisotto, & Dunn, 2012; Murphy, 2007), by promoting a more engaging interpersonal style that might improve both the clarity of expression and perceivers' cue detection (Human et al., 2012).

The high self-presentational demands in aspects of Assessment Centre exercises (ACs; McFarland, Yun, Harold, Viera, & Moore, 2005) may therefore be strengths of this approach. That is, ACs may have good predictive validity partly *because* they give candidates opportunities to self-present, showcasing their best qualities in as best they can, capturing assessors' attention and enabling them to form more accurate judgments. If so, personality perception researchers may wish to utilize aspects of ACs more frequently, as they might yield high-quality target stimuli.

Another important factor in expressive accuracy is psychological well-being, which is a consistently strong predictor of expressive accuracy in a range of contexts, including with new acquaintances (Human & Biesanz, 2011; Human, Biesanz, Finseth, Pierce, & Le, 2014), trained observers, and close others (Colvin, 1993a, 1993b). Well-being

appears partly to foster expressive accuracy by promoting provision of relevant cues to one's broader personality traits (Human et al., 2014). Greater trait and possibly even state well-being may therefore enhance candidate expressive accuracy in selection contexts as well. Consistent with Lievens's argument for more focus on transactive processes, higher state well-being could be facilitated by assessors or judges. Indeed, having a good judge in an interaction tends to increase other judges' accuracy, potentially because good judges put targets at ease (Letzring, 2008). Thus, increasing candidate well-being in selection contexts, through assessor behaviour (e.g. warmth and friendliness) or other means, could contribute to identifying the right candidate.

Expressive accuracy may not only be sensitive to contextual factors but may also be learned and developed through exposure and training (e.g. Halberstadt, 1983; see Human & Biesanz, 2013 for review). Training targets may therefore be fruitful in enhancing the accuracy of assessor judgments. Combining assessor perceptive accuracy training (perhaps through dispositional reasoning, as Lievens suggested) with candidate expressive accuracy training may in fact be necessary, as it appears that being a good judge may not be that helpful unless one is judging a sufficiently good target (Rogers & Biesanz, 2016; Zaki, Bolger, & Ochsner, 2008). This is consistent with the multiplicative nature of the realistic accuracy model (RAM; Funder, 1995), which posits that success of the latter stages (i.e. cue detection and utilization) depends on successful achievement of prior stages (i.e. cue availability and relevance).

That said, training candidates to be expressively accurate might at first seem less feasible than training assessors because candidates may not share the same motivation for accuracy; whereas assessors may usually wish to be accurate to make the right choice, candidates may seek to conceal certain flaws. But being perceived accurately, flaws and all,

could have benefits for candidates. Indeed, people tend to prefer those they accurately perceive, even if that might involve perceiving some less positive characteristics (e.g. Human, Sandstrom, Biesanz, & Dunn, 2013) This may be because people enjoy predictable individuals who confirm their expectations (Swann, Johnson, & Bosson, 2009) and who are easier to process or understand (e.g. Reber, Schwarz, Winkielman, 2004). This may be particularly true for high-stakes contexts such as personnel selection, where assessors may prefer candidates whom they accurately perceive, despite awareness of some flaws, to taking chances on candidates whom they cannot quite read. In fact, Moore, Lee, Kim, & Cable (2017) found that candidates' honesty in an interview enhanced their hire ability, only if they were high-quality candidates. Therefore, fostering expressive accuracy may increase assessors' ability to identify high-quality, suitable candidates for the job and may also enhance (high-quality) candidates' likelihood of being hired.

CONCLUSION

In sum, personnel selection and personality perception research can certainly benefit from each other's approaches and findings. We agree that ACs, dispositional reasoning, and transactive processes deserve greater attention. However, personnel selection (and personality perception) research would benefit from greater focus on expressive accuracy. Specifically, it may be possible to harness insights on predictors of the good target to enhance accuracy of judgments in personnel selection contexts. This may enable assessors to select the best candidates for appropriate positions and also allow candidates to reap the benefits of being good targets.

Implicit Trait Policies in Personality Research

STEPHAN J. MOTOWIDLO

Rice University, USA
stephan.j.motowidlo@rice.edu

Abstract: I elaborate Lievens's suggestion that personality researchers may find it useful to study implicit trait policies (ITPs), which are implicit beliefs about correlations between personality traits and behavioural effectiveness. I emphasize that an ITP is not the same as a personality trait. Instead, personality traits are antecedents of ITPs. Based upon our model which positions ITPs as mediators of effects of personality traits on behaviour, I offer the hypothesis that personality influences goal-oriented behaviour only through its effects on ITPs. Copyright © 2017 European Association of Personality Psychology

I was delighted to read Lievens's article on possibilities for interplay between personality and personnel selection research. He made many excellent points and presented a compelling case that ideas developed by selection researchers about personality expression can contribute to personality research more generally. Although his discussion covered several interesting and important topics related to situational judgment tests (SJTs) and assessment centre

exercises, I focus on only one point that he developed in connection with SJTs. He mentioned the concept of implicit trait policy (ITP) and discussed its role in SJT research and its potential implications for personality research. I underscore and elaborate upon some of the points he made about it.

We developed the concept of ITP (Motowidlo, Hooper & Jackson, 2006) to explain why SJT scores were correlated with personality traits such as agreeableness. We suggested

that (a) since many SJTs have interpersonal content, perhaps response options recognized by the scoring key as most effective often express agreeableness and (b) perhaps agreeable people are more likely to believe that agreeable action is more effective than disagreeable action. If so, agreeable people would correctly identify the agreeable SJT response options as the most effective ones or as ones they would be most likely to carry out. So ITP is a belief about how much more effective specific trait-related actions are than less agreeable or disagreeable actions. And people high on the related traits are more likely than those lower to believe that trait-related actions are more effective.

This last point highlights the difference between a trait such as agreeableness and an ITP that targets that trait. Our theory proposes that people who are, say, more agreeable are more likely to develop ITPs that value agreeable actions. Consequently, agreeableness is an antecedent of ITP for agreeableness. But it is not the only antecedent. We posit that people learn through experience about the relative value of agreeable and disagreeable action, so this learning is another antecedent. And this, of course, means that an ITP involving agreeableness is not the same as the basic trait of agreeableness.

ITPs measured with SJTs are interesting to selection researchers largely because of their potential for predicting job performance. Lievens insightfully noted that for selection researchers, job effectiveness is the important goal that drives interest in personality. SJTs implicitly allude to this goal when they ask about the effectiveness of response options. Standard self-report measures of personality do not typically refer to any particular goals when asking respondents to describe themselves according to statements that reflect one trait or another. So if ITPs measured by SJTs are used by personality researchers, they will probably need to be couched in the framework of goals, implicitly or explicitly. Lievens mentioned that 'there is no reason that SJT items should be work-related. Items with situations related to school, health, romantic relationships, sports, etc. could also be developed'. I agree, but I would add that even if not work-related, SJT items still hinge on some goal or desirable state to be achieved by executing one SJT response option or another. If we ask respondents to pick the best or most effective action from a set of 4 or 5 alternative response options about different ways to handle a romantic crisis, there must be some goal, even if implicit and unstated, to be achieved by performing the action. For instance, if the actions vary in the level of agreeableness they express, we can score responses to yield a measure of ITP for agreeableness with

respect to romantic relations. We might discover that some people believe that the key to success in romantic relations is expressing agreeableness, while other people believe the key is in expressing dominance or authority. We would expect to see people act accordingly (with agreeableness or domination) when in romantic situations.

But this relies on assumption that beliefs about the best way to handle, say, romantic relations is partly learned and partly influenced by basic traits for agreeableness or dominance. We do not expect perfect or even particularly strong relations between someone's basic trait of agreeableness and their ITP for agreeableness in romantic situations. This is an important point because if personality researchers become interested in studying ITPs measured through SJTs, they should not expect to be measuring personality constructs. They should instead think about relations between the ITP of interest and its antecedents in basic personality traits and learning from experience.

I suspect the notion of ITP will be most interesting to personality researchers concerned with predicting behavioural patterns that are innately functional or dysfunctional for individuals or society at large. In that case, their focus would be on the goal represented by these desirable or undesirable behavioural patterns. As Lievens pointed out, many studies have shown that SJT scores, which presumably are laced, deliberately or unintentionally, with ITPs for one trait or another, predict behavioural criterion variance beyond that predicted by personality tests. This fits well with our theory that proposes that ITPs mediate relations between personality traits and behaviour.

In sum, I have argued that ITPs can be useful objects of study for personality researchers, especially those interested in predicting or explaining behavioural patterns that are intrinsically functional or dysfunctional for individuals or society. I have also pointed out that ITP scores derived from SJTs should not be regarded as equivalent to personality scores, but they are interesting and important anyway because of their potential for predicting patterns of behaviour that we care about. Their connection with personality derives from their theoretical role, at least in our model, as the path through which personality indirectly drives important behaviour. I close by offering a specific hypothesis which I hope will invite efforts at disconfirmation: basic personality traits such as those represented by the Big Five affect goal-oriented behaviour *only* through their effects on ITPs, which are beliefs about the connection between expressing those traits and achieving some goal.

We Can Do More with Already-Existing 'Tricks'

RENÉ MÕTTUS[1,2]

[1]University of Edinburgh, Edinburgh, UK
[2]University of Tartu, Tartu, Estonia
rene.mottus@ed.ac.uk

Abstract: To increase predictive accuracy at no extra cost, life outcomes including job performance can be predicted from individual questionnaire items in addition to or even instead of composite trait scores. Items can be conceived as markers of 'persome' (the universe of behavioural, affective, cognitive and motivational variance

among people), like individual genetic variants track genomic variance. Ideally, comprehensive item pools are used to capture persomic variance, but items of existing questionnaires can also provide incremental predictive accuracy. Item-level associations may reveal that outcomes are highly poly-causal. Copyright © 2017 European Association of Personality Psychology

Individuals—focally, job applicants—differ in how well they perform given jobs, and this is consequential for employers. Those involved in personnel selection therefore have the goal of working out the best possible methods to predict these performance differences from other, potentially readily observable behavioural, cognitive, affective or motivational characteristics. This partly overlaps with what interests many personality researchers. They see that individuals vary in all sorts of life-course aspects (outcomes), and they, too, want to work out how to predict these differences optimally from a multitude of potentially readily observable behavioural, cognitive, affective or motivational characteristics. With this, they hope to learn about how and the extents to which personality plays out in everyday life.

But we cannot measure everything. Personality researchers have therefore invested substantial efforts in systematizing these potentially observable characteristics, reducing them to ever fewer, broader, more abstract traits such as the Big Five. Ostensibly, these broad traits underlie observable characteristics and are thereby also causal forces partly driving the outcomes. The Big Five, measured using item composites, has been correlated with hundreds of outcomes, including those that are of interest in personnel selection.

My reading of Lievens's article is that personnel selection researchers say: 'broad traits such as the Big Five are fine, but we think that we could actually be doing even better'. In part, this is due to limitations of the self-report method. But in part, personnel selection could do better by capturing individual differences in more nuanced and context-sensitive ways. And guess what, I think personality researchers may have reasons for doing this too. Broad trait constructs have served us well, but by now, we may have learned most of what they can teach us.

Drawing a loose parallel with genome, let's call all the possible behavioural, affective, cognitive and motivational differences among people the 'persome'. In both personality research and personnel selection, we need as broad coverage of the persome as possible. Broad dispositional traits capture some of the persomic variance, but they also miss much because they do not encompass all potentially relevant ways individuals vary. Furthermore, they may misrepresent how personality is linked with outcomes because it may not be the underlying structures of the persome personality psychologists think the broad traits ought to reflect that drive their associations with outcomes but specific constituents of the aggregates (Mõttus, 2016).

Genomic variance is captured by tracking individual genetic variants scattered throughout the genome. Perhaps personality research and personnel selection could do something similar, by tracking markers of persomic variance scattered as thoroughly as possible throughout the persome. The markers can be operationalized as individual questionnaire items, reflecting specific and more context-sensitive personality characteristics that have been called nuances

(McCrae, 2015, Mõttus, Kandler, Bleidorn, Riemann & McCrae, 2017a). This would capture maximal persomic variance. It would mean sampling the persome, as we do sample populations. Thereby, relating these markers (individual items) to whatever we want to predict—job performance, antisocial behaviour or longevity—may offer more accurate predictions than those based on broad composites of only a few markers (particularly those that allegedly 'best' measure broad traits such as the Big Five).

Yes, measurement error is likely greater in single items than in their aggregates, and the more parameters in models, the more prone they are to over-fitting. For this reason, predictive models should be created (or trained, using machine learning terminology) using large samples and tested in other representative samples. This is another parallel with genetic research: as the training samples (i.e. genome-wide association studies; GWAS) are getting bigger (from a few to hundreds of thousands), an increasing amount of variance is being predicted from genomic data in independent samples (Cesarini & Visscher, 2017).

And yes, we may worry about apparent loss of parsimony: isn't it nice to see a few familiar composite traits correlate with whatever we want to predict with personality? One may feel less comfortable with the idea that predictive models of outcomes require hundreds of predictors with commensurately small associations. But, again, if lessons from genetic research are relevant, this may reflect reality: just as complex traits are highly polygenic (Chabris, Lee, Cesarini, Benjamin, & Laibson, 2015), they may also be highly poly-causal insomuch as personality is concerned.

Quite likely, training item-based prediction models of outcomes, including those relevant in personnel selection, would be an atheoretical enterprise, similar to how GWAS atheoretically link genetic variants and observable characteristics. Scary? Maybe but there are arguments that empiricism should trump a priori implicit assumptions (Yarkoni & Westfall, in press). For example, many hypotheses—unless they are trivial—tend to be wrong (for a parallel, think of candidate gene research).

To address the possibility that outcomes are linked with large numbers of very specific characteristics, we need comprehensive item pools rather than item-sets of questionnaires that have been designed to measure particular traits such as the Big Five. However, even already existing data collected with these less-than-ideally comprehensive item pools are consistent with such a poly-causal world. For example, we observed that item-based predictive models almost always outperformed those based on the Big Five or their facets in predicting a range of life outcomes in independent samples (Mõttus, Bates, Condon, Mroczek, & Revelle, 2017b). Although job performance was not among these outcomes, it is very likely that predicting it from individual questionnaire items rather than the Big Five or its facets would generate more accurate predictions. Importantly, this leverage comes

at no extra cost as one needs to have item data to calculate trait scores anyway.

Therefore, in addition to, or perhaps even before, developing new methods to predict job performance or any other life outcome from personality, it may make sense to make the best possible use of already-existing 'tricks' and already-existing data—in personality psychology and personnel selection alike.

Assessing Personality Traits in Specific Situations: What Situational Judgment Tests Can and Cannot Do

PATRICK MUSSEL, PHILIPP SCHÄPERS, JAN-PHILIPP SCHULZ, JULIAN SCHULZE and STEFAN KRUMM

Freie Universität, Berlin, Germany
patrick.mussel@fu-berlin.de

Abstract: We echo Lievens's call to incorporate measures developed in applied settings as valuable instruments for personality researchers. We highlight some striking findings that have challenged central assumptions about these measures and call for joint research efforts of personality and industrial and organizational psychologists to gain better understanding of what these methods actually measure. Copyright © 2017 European Association of Personality Psychology

Personnel selection is a major task for organizations in the public and private sectors. Accordingly, industrial and organizational psychologists have developed methods to accomplish this challenging undertaking by utilizing and improving existing diagnostic methods and by creating new methods for the purposes of predicting job performance and related criteria. As these purposes relate to assessment in specific work-related domains, these methods are well suited to assess interplay of dispositional traits with situational characteristics. Lievens's article makes a significant contribution by outlining how two of these methods, Situational Judgment Tests (SJTs) and Assessment Centre exercises (ACs), might be successfully utilized to address important research and applied questions, including within-person variability, personality disorders, trait-behaviour links, and personality expression and perception. Using these themes, he developed an impressive and highly interesting research agenda that could probably keep a whole cohort of PhD students busy for a while!

Methods for assessing personality traits have several weaknesses, such as being susceptible to response distortion or requiring, to a certain degree, introspective ability (Ziegler, MacCann & Roberts, 2011). However, at least at first sight, it seems questionable whether methods from the industrial and organizational domain are suited to remedy these weaknesses. Developed with the goal of maximizing criterion-related validity, these methods sometimes have other psychometric problems, such as reliability or construct validity. Dimension ratings from ACs or employment interviews, for example, lack convergent and discriminant validity, despite showing significant relations with work-related criteria (Lievens & Conway, 2001; Mussel, 2007; Sackett & Dreher, 1982). Similarly, SJTs often lack construct-related validity (Christian, Edwards, & Bradley, 2010). Mussel, Gatzka and Hewig (2016) developed a Big Five SJT with the goal of addressing this shortcoming, thereby modifying traditional SJTs in four regards: first, we aimed to assess narrow facets, rather than broad traits or compound variables. Second, we used well-defined factors from established

personality models, rather than vague or diffuse competence profiles. Third, response options were developed and scored to reflect high versus low levels of traits, rather than high versus low effectiveness. Fourth, we developed a rather large number of items (22) per trait.

As Lievens outlined, this particular SJT showed high levels of convergent and discriminant validity. While these results are promising, future research is necessary to identify the extent to which each of the above-mentioned features contributed towards this goal. Additionally, it is necessary to show that these modifications did not negatively affect other appreciated properties of SJTs, such as applicant reaction or criterion-related validity. Such an effect occurred, for example, in a study by Kleinmann (1997), who found that making AC dimensions transparent to candidates improved construct-validity but at the same time negatively affected criterion-related validity.

As Lievens also noted, a distinctive feature of SJTs is contextualization, that is assessment of behavioural intentions or effectiveness ratings regarding specific situations. Given this, SJTs are assumed to be measures of context-dependent knowledge, where the situation descriptions play key roles: participants need to be able to envision the situations and make judgments about alternative ways of responding that mirror the judgments they would make in similar situations in reality (Campion & Ployhart, 2013). Interestingly, this assumption was challenged by Krumm et al. (2015), who compared traditional SJT ratings with a modified version which completely lacked situational descriptions, thus consisting solely of response options. Provision of contextual information usually provided in the situational description had strikingly little effect: across two studies, for 43% to 71% of the items, it did not matter whether situation descriptions were included or not to the number of correct responses per item. Thus, an important task for future studies is to examine whether the behavioural intentions expressed in SJT responses reflect situation-depended real behaviour, rather than more generic dispositions. Stated otherwise, how much variance in SJTs is due to situational characteristics?

Incorporating situational taxonomies such as the DIA-MONDS (Rauthmann et al., 2014) might be a promising avenue in this regard, even though they provide operationalizations of situation–person interactions (rather than situation effects per se) and, thus, would not necessarily lead to better discrimination of within- and between-person variance.

Another topic that might be worth discussing is the role of implicit trait policies (ITP) as theoretical background for explaining what SJTs assess. The definition of ITPs 'as 'implicit beliefs about causal relations between personality traits and behavioural effectiveness' (Motowidlo, Hooper & Jackson, 2006, p. 749) is strongly tied to its operationalization, which is a special technique relating effectiveness ratings of response options with expert ratings regarding extents to which these response options reflect certain levels of given traits. ITP ratings, thus, can be seen as implicit personality assessments, similar to objective personality tests. As Motowidlo et al. (2006) noted, ITPs require that response options that represent high levels of given traits are truly more effective, an assumption that is not readily fulfilled for SJTs with effectiveness ratings. Additionally, there still seems to be some disagreement as how to conceptualize ITPs, as they are inconsistently referred to as policies, beliefs, procedural knowledge, general domain knowledge, or personality traits (Motowidlo et al., 2006; Lievens & Motowidlo, 2016). Finally, while SJTs are typically used to gauge ITPs, empirical evidence is needed to show that ITPs, derived from one SJT, predict performance in another, traditionally scored SJT.

In sum, we echo Filip Lievens's call to incorporate measures such as SJTs and ACs as fruitful alternative assessment methods for personality researchers and at the same time call for research that helps to understand these methods better, including their benefits and limitations. On a related note, industrial and organizational psychology could likewise benefit from acquisitions from personality psychology, for example by focusing on well-defined constructs instead of fuzzy compound variables.

Cautions About the Usefulness of Indirect Measures of Personality Based on Personnel Selection Methods

IN-SUE OH[1], YOU JIN KIM[1] and MYEONG JIN KIM[2]

[1]Temple University, Philadelphia, PA, USA
[2]Yonsei University, Seoul, Korea
insue.oh@temple.edu

Abstract: Lievens discussed how two personnel selection methods, situational judgment tests and assessment centre exercises, can be used as potentially useful, indirect measurement methods for personality, thus helping to build a bridge between two separate streams of research on personality assessment and personnel selection and ultimately improving the ecological and predictive validity of personality. We provide some alternative perspectives regarding several assumptions and cost-effectiveness issues underlying Lievens's arguments and suggestions. Copyright © 2017 European Association of Personality Psychology

Lievens encouraged personality researchers to develop and validate situational judgment test (SJT) items that assess various personality traits. However, we believe that SJTs are not suited for measuring some personality traits (in cost-effective ways). We begin by discussing some important measurement properties of SJTs. SJTs instruct respondents to put themselves in specified sets of hypothetical situations (items) and ask how they would behave in those situations. This implicitly requires assumption that individual differences in behavioural responses to those pre-specified situations will generalize to other situations and/or that those pre-specified situations are representative samples of all possible relevant situations, from which we can sufficiently infer individual differences in personality. Measuring personality this way is based on another assumption that individuals choose to enter the same situations in the same sequence (given the pre-determined array of SJT items). However, these assumptions are difficult to justify for some personality traits.

Bateman and Crant (1993) argued that proactive individuals choose their environments through selection. Proactive individuals purposely search for best-fitting environments via boundary scanning so that they spend greater effort choosing environments that allow them to manifest their proactive personalities behaviorally (e.g. initiate changes). Furthermore, proactive individuals do not limit their behavioural reactions to a situation to pre-specified options but often choose to change the situation in question. Therefore, use of response options in SJT items limits the ecological validity of assessing the totality of proactive personality. Thus, if we want to use SJTs to measure proactive personality, as Lievens noted, researchers need to adopt nonlinear SJTs in which respondents' answers to earlier items influence the subsequent situations (items) that they receive. Alternatively, researchers can use assessment centre (AC) exercises, because most ACs do not force respondents to enter into the same situations in the same sequence and allow them sufficient autonomy to change their situations.

In addition, SJTs and/or ACs may be less effective in measuring dark-side personality traits for two reasons: (a) these negative traits are more subject to response distortion (given their obvious negativity) in high-stakes settings and not easily detectable nor visible in typical work settings. Thus, it is difficult to come up with SJT items measuring these traits, and these negative traits are unlikely to be activated in strong situations like ACs where individuals are asked to interact with others and (they are aware that) their behaviours are being actively evaluated by assessors for personnel purposes.

We believe that conditional reasoning tests (CRTs), originally developed by industrial and organizational (I/O) psychologists, can be used instead to measure some negative personality traits such as aggression (James, 1998). Similar to SJTs, CRTs are also indirect measures of personality designed to infer personality from individual differences in reasoning processes (e.g. various justification mechanisms such as attribution), based on the assumption that personality also affects how we reason and solve problems (James, 1998). The most widely used CRT is the CRT for Aggression (CRT-A), which presents a brief statement and asks respondents to choose one option they think most appropriate for interpretating the statement from four possible options: an option that seems logical or reasonable for those with hostile orientation, an option that seems logical or reasonable for those with prosocial orientation, and two other options that are completely unreasonable (James & McIntyre, 2000). James (1998) argued that aggressive individuals are more likely to choose the first option as the most appropriate and later engage in aggressive behaviour in similar real-life situations because the reasoning processes underlying their thoughts and actions are hostilely oriented. Existing meta-analytic evidence indicates that the CRT-A is as predictive of job performance as other relevant self-report measures of personality (Berry, Sackett, & Tobares, 2010). In addition, on the surface, the CRT looks more like a reasoning test than a personality test, and, thus, it is perhaps less subject to faking and perceived to be more favourable among job applicants than corresponding self-report measures of personality.

Although personnel selection-based measurement of personality has merit, it is more costly and difficult to develop (and administer) such indirect measures of personality than direct measures of personality. This is even more true if we want to develop full-spectrum measures of personality based on the Big Five or HEXACO and their equivalent measures. Relatedly, recent evidence in I/O psychology is clear that informant ratings of personality are more valid than self-reports of personality and provide meaningful incremental validity over self-reports of personality in predicting important academic and occupational outcomes (e.g. Oh, Wang, & Mount, 2011). This evidence is impressive given that informant ratings of personality are sufficiently (more highly) correlated with self-reports of personality (than other personnel selection methods-based personality ratings; Connelly & Ones, 2010). Currently, we have no evidence that indirect measures of personality are more valid than informant ratings of personality and offer meaningful incremental validity over informant ratings of personality for important outcomes. As such, the utility of indirect measures of personality vis-à-vis information ratings of personality is unclear. Finally, indirect measures of personality require high levels of reasoning skills and, thus, may measure more than the intended personality traits such as cognitive ability, thereby generating unintended subgroup differences (an important issue in personnel selection). Thus, we also need evidence that indirect measures of personality are as non-cognitive as direct measures of personality.

In conclusion, we certainly agree that personnel selection methods (e.g. SJTs) can be used together with direct measures of personality in integrative manners to cross-fertilize two separate streams of research on personality assessment and personnel selection through application of multitrait-multimethod and ultimately improve the ecological and predictive validity of personality. However, given much higher costs involved in item development (and administration) processes, we need more evidence before we accept that personnel selection methods-based indirect measures of personality can be more useful than direct personality measures (particularly, informant ratings).

Validity Risks and Potential Advancements of Situational Judgment Tests and Assessment Centre Exercises in Personality Research

JOHANNA PRETSCH and MANFRED SCHMITT

University of Koblenz-Landau, Landau, Germany
schmittm@uni-landau.de

Abstract: Lievens proposed use of Situational Judgment Tests (SJTs) and Assessment Centre exercises (ACs) for personality research. We suggest potential limitations of these methods and offer idea for improving them. We propose that SJT can be enriched by considering the equi- and multi-finality of trait expressions, multiple goals, and the basic assumption of rational choice theory. We argue that potential threats to the validity of ACs in personality assessment can result from hierarchies of trait visibility in ACs and trait expression dynamics in group activities. Copyright © 2017 European Association of Personality Psychology

Lievens proposed the use of two methods for personality research: Situational Judgment Tests (SJTs) and Assessment Centre exercises (ACs). We suggest potential limitations of these methods and offer ideas for improving them.

ENRICHING SJTS BY CONSIDERING THE EQUI-FINALITY AND MULTI-FINALITY OF TRAIT EXPRESSIONS AND MULTIPLE GOALS

The way SJT and Implicit Trait Policy (ITP) measures are designed (see appendix 1 in Lievens's article) limits their diagnostic potential for two reasons. The first follows from the principles of behavioural multi- and equi-finality (Wood, Gardner, & Harms, 2015). Multi-finality occurs when specific behaviours have multiple consequences. For instance, not only is working on a task instrumental for completing the task but it also incurs opportunity costs concerning other tasks or projects. Equi-finality occurs when different behaviours can have identical consequences and thus be equally instrumental. Support from a colleague can be obtained by asking for help, making a deal, or appealing to organizational norms. Each of the equi-final behaviours may express a different trait. The second reason follows from Expectancy × Value Theory (Fishbein & Ajzen, 1975) and Rational Choice Theory (Coleman & Fararo, 1992). According to these theories, behavioural choices are made by jointly considering the instrumentalities of alternative behaviours and the values attached to their consequences. SJT and ITP measures are limited because they do not systematically combine behavioural alternatives (several expressions of several traits) with several possible consequences. Rather, the goal is implicitly assumed in the description of the situation. For example, the description of a conflict implicitly assumes the goal to resolve it quickly. However, doing this might not be the only reasonable goal in the situation. For example, a person might believe that escalating the conflict now could lead to more beneficial long-run resolution.

We suggest that tapping several traits and offering several potential goals explicitly in an SJT would provide additional diagnostic information. This would make SJTs better operationalizations of established behavioural decision theories such as Expectancy × Value and Rational Choice Theories. Crossing several traits with several goals in SJT items would enable a fine-grained picture of how individuals express traits when faced with several competing goals and values. It would also aid in assessing the relative flexibility of trait expressions in the complex situations typical of many life contexts. This would contribute to a more comprehensive, realistic, and externally valid assessment of behavioural variability within individuals. Finally, a variant of the proposed extension might be useful for assessing individuals' social competence: instead of exposing participants to a fixed set of goals, they could be asked to articulate goals in specific situations. The number of goals and complexity of links drawn between these goals and behavioural choices might be good indicators of social and technical competence.

POTENTIAL THREATS TO THE VALIDITY OF AC EXERCISES IN PERSONALITY ASSESSMENT AND RESEARCH

According to Trait Activation Theory (Tett & Guterman, 2000), ACs offer particular situational cues that lead people to display trait-relevant behaviour. In the AC literature and practice, ACs are always considered to provide activation cues for several traits. However, to our knowledge, it is not clear whether several simultaneously activated traits can be identified by assessors. Therefore, we wonder whether there is a *hierarchy of trait visibility* for the traits activated in an exercise. There could be several reasons for such a hierarchy. First could be relative expression visibility (e.g. behavioural expression of Conscientiousness might impede detection of Extraversion expression, which is also activated). The resulting hierarchy might not vary among individuals and might instead depend on characteristics of the specific trait-relevant behaviours. Second could be relative extent of expression of one or several traits that are intended to be activated in an exercise. Participants might use only some of the situational cues and therefore express only some traits or express some traits more than others. A hierarchy of trait visibility resulting from differential cue utilization would thus likely vary among individuals. Either hierarchy could have major implications for reliability and validity of trait assessments in ACs. First, separate exercises cannot be regarded as parallel tests. Second, absolute levels of less visible traits could be underestimated. Third, observers might have trouble detecting differences among individuals in traits lower in the hierarchy, potentially manifesting in lower inter-assessor agreement or the like.

Many ACs are designed as group activities (e.g. group role-plays among competing candidates). Here, individuals' trait expressions will depend on their fellow candidates' trait expressions, and each will adapt behaviour to the others' behaviours. It could thus be that some candidates' authentic personality-congruent behaviour would not be optimal in a given situation because of the group's 'trait composition'. For example, consider two broad behavioural categories, agency and communion: say Candidate A jumps in to act constructively agentically in the situation. Candidate B might then see that, though he/she often prefers agentic roles, the most effective role now is communal support for Candidate A's actions. Whereas it is generally suggested that personality-non-congruent behaviour is costly (e.g. Sheldon & Kasser, 1995), competing using the same behavioural strategies for a limited resource is often even more costly. This idea is consistent with a general view of personality variation as distribution of viable strategies (MacDonald, 1995). These considerations may be especially relevant when restriction of trait variance in participating individuals is likely. Restricted variance should increase the likelihood of personality-non-congruent behaviours, thus bringing along validity risks for ACs. Absolute trait levels could be underestimated, and differences among individuals might be overestimated.

Using Situations as Diagnostic Agents for Personality Assessment

JOHN F. RAUTHMANN

Wake Forest University, USA
jfrauthmann@gmail.com

Abstract: Lievens presented suggestions for integrating assessment centre exercises into personality science, such as developing exercises based on situational taxonomies to examine trait activation potentials. To reach this goal, I argue that we can use situations, specifically their psychological content characteristics (e.g. DIAMONDS) and style characteristics (e.g. situational strength), as diagnostic agents for detecting individual differences among job candidates, existing personnel, or leaders that are meaningful or important for specific tasks, occupations, teams, work settings, or organizations. Copyright © 2017 European Association of Personality Psychology

Lievens delineated suggestions for integrating assessment centre exercises (ACs) into personality science, with the hope to 'instigate many fruitful cross-disciplinary investigations and collaborations'. I wholeheartedly agree that personality science can tremendously benefit from engaging more multi-disciplinary projects (e.g. in consortia). Additionally, I specifically welcome that Lievens sketched how applied and personality psychology can cross-fertilize. Here, I focus on one of the domains Lievens noted needs more attention: developing assessment centre exercises based on situational taxonomies to examine trait-activation potentials. This goal presupposes that we can use situations, specifically their *content* and *style characteristics* (Rauthmann, 2015), as diagnostic agents for detecting individual differences among job candidates, existing personnel, or leaders that are meaningful or important for given tasks, occupations, teams, work settings, or organizations.

SITUATION CHARACTERISTICS AS DIAGNOSTIC AGENTS

Situation characteristics describe what a situation is psychologically about (Rauthmann, 2015), and various taxonomies have been proposed to organize them (Horstmann, Rauthmann, & Sherman, 2017). One example is the Situational Eight DIAMONDS taxonomy (Rauthmann et al., 2014; Rauthmann & Sherman, 2016a, 2016b), with the dimensions Duty (*work* and *tasks*), Intellect (*processing* and *problem-solving*), Adversity (*criticism* and *domineering*), Mating (*sex* and *romance*), pOsitivity (*fun* and *playfulness*), Negativity (*stress* and *frustration*), Deception (*mistrust* and *sabotage*), and Sociality (*relations* and *cooperation*). ACs could incorporate situation characteristics, such as the DIAMONDS, in one of (at least) two ways: using situations as means or as outcomes.

First, ACs could 'build' settings where a particular DIAMONDS dimension (or several for more complex situations) is supposed to be salient to activate diagnostically interesting individual differences in trait-relevant behaviour. Because each DIAMONDS dimension is uniquely associated with different behaviours relevant to Big Five or HEXACO traits (Rauthmann, Sherman, Nave, et al., 2015; Rauthmann, Jones, & Sherman, 2016; Sherman et al., 2015), situations

with certain DIAMONDS potentials could reveal different trait-relevant behaviours (Table 1). Thus, it is important to attend to *situational content relevance*: is a given AC situation relevant to (and activating) behaviours we want or need to observe? AC situations could be constructed with great care to activate only those DIAMONDS dimensions intended. Towards this end, several test rounds with raters *in situ* (AC participants), *juxta situm* (on-site assessors), and *ex situ* (coders using videos of the AC situation) should judge an AC situation until, after optimization rounds, it taps the intended characteristics. Here, situations would be used as means towards ends: being able to assess individual differences in behaviour supposedly diagnostic of traits validly and reliably, especially when situations are maximally relevant, repeated, or prolonged.

Second, situations could also be used as outcomes that are explicitly assessed for characteristics and important in their own rights. AC assessors may observe, score, and evaluate how assesses navigate given AC settings and what they 'do' to the situations. For example, to what extent and how people changed, evoked, or created new situations in AC settings may be diagnostically relevant. Assessors would be attending not only to assesses' behaviours *in situ* (as is already traditionally done; see above) but also to *changes* in situations' content characteristics as well as assesses' situation management strategies to bring about such change (Rauthmann, 2017; Rauthmann & Sherman, 2016c). It is thus important to quantify (a) to what extent a situation has changed as a function of a given assessee's behaviour (e.g. did the situation become more intellectually stimulating, adverse, etc.?) and (b) why the situation changed (e.g. did the assesse unwillingly evoke reactions from others, purposefully change the situation, etc.?).

SITUATION STYLE CHARACTERISTICS AS DIAGNOSTIC AGENTS

Style characteristics lack content and describe how situations work or operate (Rauthmann, 2015), such as *typical*, *ordinary*, *normal*, *regular*, *usual, sudden*, *expected*, *earlier*, *late(r)*, *old*, *new*, *known*, *static*, *dynamic*, *periodic*, *unique*, *specific*, and *immediate*. One pivotal style characteristic is *situational strength* (Cooper & Withey, 2009; Dalal et al., 2015;

Table 1. The relevance of situation characteristics to diagnostically relevant information

Situation characteristic	Activated trait behaviour	Diagnostically interesting questions
Duty	Conscientiousness	• Will the person do the job? • Can the person work hard?
Intellect	Openness/intellect	• Will the person put thought into the work? • Will the person successfully solve problems?
Adversity	(Dis)Agreeableness Emotional stability	• How does the person handle criticism? • Can the person be tough if need be?
Mating	Extraversion	(Usually not relevant in AC settings)[a]
pOsitivity	Extraversion	• Can the person create a positive working environment? • Is the person not distracted by playful elements?
Negativity	Neuroticism	• Can the person cope with stress? • How does the person react to frustrations?
Deception	Honesty/humility	• Does the person uphold ethical principles and standards? • Can the person build a trusting, open, and transparent working environment?
Sociality	Agreeableness Extraversion	• Can the person establish good relations with team members? • Can the person cooperate?

Note: Coordination of situation characteristics and behaviours was done according to Rauthmann et al. (2014, 2016), Rauthmann, Sherman, Nave, et al. (2015) and Sherman et al. (2015).

[a]AC = Assessment centre. The mating dimension does not readily lend itself to evaluative performance assessment in work settings. While sexual and romantic relationships may occur during work, mating characteristics of situations in ACs are likely often not of relevance to hiring, promoting, or developing personnel. Nonetheless, in certain circumstances, such as sexual harassment and gender sensitivity training, mating potential of situations may be of interest.

Table 2. Situation strength matrix

Within versus across situations	Idiographic for persons: *Within one person*		Nomothetic for persons: *Across different persons*	
	Behavioural options: Are behaviours enacted or not?	Behavioural level: How strongly is a given behaviour expressed?	Behavioural options: Are behaviours enacted or not?	Behavioural level: How strongly is a given behaviour expressed?
Idiographic for Situations: *Within one strong situation*	Behavioural options are diminished: One person can only select a small subset of behaviours from her/his repertoire	Intensity of behaviour expression is diminished or increased: The person expresses some behaviours strongly, others weakly	Between-person variation in which behaviours are shown is diminished: Most persons select similar behaviour(s)	Between-person variation in intensity of behaviour expression is diminished or increased: Most persons express some behaviours strongly, others weakly
Nomothetic for Situations: *Across several instances of strong situations*	Cross-occasion within-person variation in behaviour selection is diminished: The person always chooses the same behaviour	Cross-occasion within-person variation in the intensity-expression of a behaviour is diminished: The person always expresses a given behaviour in similar intensity	Between-person variation in cross-occasion within-person variation in behaviour selection is diminished: Most persons choose the same behaviour again and again	Between-person variation in cross-occasion within-person variation in intensity of behaviour expression is diminished: Most persons express a given behaviour with similar intensity again and again

Meyer & Dalal, 2009; Meyer, Dalal, & Hermida, 2010), which Lievens also discussed. I propose a nuanced Situation Strength Matrix (Table 2) varying person, behaviour, and situation modalities. First, how strong/weak situations operate can be examined at the within-person level for each person separately (which is more idiographic) or at the between-person level across different persons (which is more nomothetic). Second, one could either focus on which behaviours from some given total repertoire of behaviours are expressed (either Yes or No) or how strongly each behaviour is expressed (i.e. intensity of expressions). Third, one can compare these two points for one situation presumed strong/weak at one given timepoint in several instances to which participants are repeatedly exposed (e.g. in trainings). Fully crossing these options yields a $2 \times 2 \times 2 = 8$ cell matrix of different 'tests' of to what extent a situation may be considered strong. Theoretically, the more cells a given situation affects, the stronger it may be considered, although theory might predict that some cells are more important than others or not relevant at all. To my knowledge, such fine-grained distinctions have not been made so far, and the literature has just focused on one or maybe two cells at a time. Such a differentiated conceptualization of situational strength can help gauge what behavioural outcomes to expect, and AC settings should carefully choose, and be explicit about, which cell(s) of the Situation Strength Matrix they want/need to tap.

ACs could then employ situations of differing strengths, which should also be optimized for the content characteristics they tap (Table 1).

CONCLUSION

Lievens touched upon important but often neglected topics in mainstream personality psychology. I hope that researchers take up his suggestions and start conducting new and exciting lines of research. As I outlined here, I believe much could be learned about personality expression by attending to situations' content and style characteristics.

Situational Construals: Key to Assessments of Personality–Situation Interplay

THOMAS ROCKSTUHL

Nanyang Technological University, Singapore
trockstuhl@ntu.edu.sg

Abstract: Lievens provided an excellent foundation for future research concerning assessment of personality–situation interplay. Inspired by his work, I elaborate on two opportunities to apply SJTs to examine situation–trait contingencies. First, I discuss the potential of open-ended SJTs to assess respondents' situational construal. Second, I suggest that experimental manipulation of situational cues in SJTs may be very useful for expanding understanding of situation–trait contingencies. Copyright © 2017 European Association of Personality Psychology

In his brilliantly crafted article, Lievens started a constructive dialogue between selection and personality researchers. He highlighted the potential of situational judgment tests (SJTs) and assessment centre exercises (ACs) to deepen research on situation–trait contingencies. His article built an excellent bridge between contemporary personnel selection and personality research and offered a treasure chest of future research ideas.

Among his suggestions, Lievens wrote: '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 close-ended (multiple-choice) formats might be used, reliance on open-ended formats in which people formulate their own response might be fruitful'.

I agree with Lievens on the potential of SJTs to illuminate situational construal. To further research along these lines, I focus on the second part of his suggestion and elaborate on (1) potential differences between close-ended (i.e. items describing situations and presenting small numbers of specific response options for them) and open-ended SJTs (requesting free responses) for assessing situational construals and (2) need to assess situational construal alongside manipulations of situational cues in SJT items.

First, Lievens noted that one challenge in using SJTs to examine situation–trait contingencies is that situations may have less influence on SJT responses than generally assumed. This caution is based on the Krumm et al. (2015) study showing similar response patterns for SJTs with and without situational descriptions. Interestingly, my colleagues and I reported in the same issue of the *Journal of Applied Psychology* that SJTs can be used to assess situational construal and that this situational construal predicted performance outcomes over and above an assessment of how people responded to the situations. One may conclude that these studies contradict each other, so it is instructive to compare them.

Krumm and colleagues drew upon existing closed-ended SJTs and showed that stripping away the situational descriptions had little effect on the chosen response options. By contrast, Rockstuhl et al. (2015) presented respondents only with situations, followed by open-ended questions about construal of the situations and likely courses of action. I argue that presenting specific response options made comparing these responses such an integral (test-)situational demand that influence of the situational descriptions was diminished. Our own open-ended approach did not cue response options, thereby giving respondents freedom to construe the situations and possible responses. Such freedom is likely key in assessing how people psychologically construe situational features in SJT items. I therefore suggest that open-ended SJTs offer particularly exciting opportunities to examine situation–trait contingencies.

Second, Lievens suggested that collaboration between selection and personality researchers in deliberately manipulating situational cues in SJT items would be fruitful. I concur wholeheartedly that this is an area where advances in situational taxonomies (Parrigon et al., 2016; Rauthman et al., 2014) are ripe for exploitation. I would only caution personality researchers to assess psychological construal of situations alongside manipulations of situational cues. What always struck me about the responses our participants offered to open-ended questions about the situation was how vastly different their situational construals of the *same* situation could be. Thus, there is likely tremendous value in exploring situation–trait contingencies not only for objective situational characteristics but also for subjective situational construals. Testing subjective situational construals alongside manipulations of situational cues in SJT items aligns well with recent

calls by personality researchers to integrate objectivist and subjectivist perspectives on situational–trait contingencies (Rauthmann, Sherman, & Funder, 2015).

All in all, Lievens laid the foundations for what can be fruitful collaborations between selection and personality researchers in examining situation–personality interplay. Undoubtedly, such collaboration will advance understanding of human behaviour and improve selection decisions in practice.

Situations in Personality Research: Offerings from the Workplace

ROBERT P. TETT

University of Tulsa, Tulsa, OK, USA
robert-tett@utulsa.edu

Abstract: Furthering Lievens's effort to extend work psychology methods to mainstream personality research, I identify key findings and principles bearing on personality in the workplace for broader application. Review of meta-analytic evidence underscores situational specificity in personality-outcome relationships, and closer examination of trait activation theory, cited by Lievens as a foundation for assessment centre design, offers personality researchers inside and outside the workplace potentially useful directions in navigating interactionist complexities.
Copyright © 2017 European Association of Personality Psychology

This comment has two parts. First, drawing from workplace meta-analyses involving personality, I advance the case for situational moderation of personality effects. Second, adopting Lievens's pluralistic aims, I describe key features of trait activation theory (Tett & Burnett, 2003; Tett, Simonet, Walser & Brown, 2013) with a view to non-work applications.

SITUATIONAL SPECIFICITY IN META-ANALYSES OF PERSONALITY-OUTCOME RELATIONS

The most frequently cited output from work-focused meta-analyses is the mean correlation (Carlson & Ji, 2011). Mean validity estimates for personality traits, as expected, vary across jobs (e.g. Salgado, 1997). Extraversion, for example, shows stronger linkages with performance in sales and managerial jobs than others (Barrick & Mount, 1991). Equally important but often ignored is substantive residual variability—not due to sampling error and other artefacts—around mean correlations with job performance. High residual variance supports situational specificity of personality validity (Hunter & Schmidt, 2004; cf. Tett, Hundley & Christiansen, 2017).

Meta-analytic evidence for situational specificity in personality-work outcome relations is substantial. The 80% credibility intervals[3] (CI) around corrected mean correlations in this area (based on artefact-corrected residual variance) average around .30 correlation units in width (Tett & Christiansen, 2007). Extraversion, for example, yields an 80% CI of −.05 to .35 within sales jobs (Barrick & Mount, 1991), indicating its importance in sales is far from uniform. Of all meta-analytic aggregations of personality-job performance validity estimates, about a third have 80% CIs with both endpoints beyond ±.10 (Tett & Christiansen, 2007).

Agreeableness seems especially prone to such 'bidirectionality', higher levels contributing to performance in some settings and lower levels in others.

Reliance on mean validity has led some (e.g. Morgeson et al., 2007) to conclude falsely that personality is, at best, weakly related to performance. In most cases, the mean is simply the midpoint of a distribution of population validities, promoting identification of conditions favouring stronger linkages. In cases of bidirectionality, mean rho severely understates the importance of personality due to cancellation of true positive and true negative values (Tett, Jackson, Rothstein & Reddon, 1999), promoting identification of conditions affecting a trait's desirability.

Prominence of situational specificity in the workplace promotes consideration of situational moderators in mainstream personality research. How personality plays out and whether it yields positive or negative outcomes can depend on complex combinations of situational features. Sorting out this complexity is a key challenge moving forward.

GENERAL OFFERINGS FROM TRAIT ACTIVATION THEORY

Lievens applied Trait Activation Theory (TAT) to assessment centre methods. Several TAT features bear extension to mainstream personality research.

First, TAT separates trait-expressive behaviour from its performance value. This is important because both positive and negative validity estimates for personality traits may derive from the same underlying processes; what differs is the judged value of the behaviour. In work settings, positive validity derives when trait expressions help meet job demands (e.g. sociability in customer service) and negative validity when expressions leave demands unmet (sociability at the water cooler). Personality expressions more broadly often have valued outcomes. Prosocial behaviour, substance abuse,

---

[3]Credibility intervals (around mean rho) are distinct from confidence intervals (around mean *r*), the latter reflecting random, second-order error in sampling *r* (Whitener, 1990; Hunter & Schmidt, 2004).

and road rage all permit conceptualization as positively or negatively valued trait expressions, and identifying situational cues that activate underlying traits offers targets for managing desirable and undesirable behaviour.

Second, TAT offers a functional taxonomy of situational features relevant to trait expression and its evaluation. Demands are cues, responses to which are positively valued; responding to distracters yields undesirable trait expressions; constraints limit availability of trait-relevant cues; releasers counteract constraints; facilitators amplify cues already present; and discretionary cues activate traits with no ties to valued outcomes, important for explaining personality effects in weak, autonomous situations. All the noted features are relevant inside and outside the workplace.

Third, TAT distinguishes among task, group, and organizational levels as sources of trait-relevant cues. Non-work analogs are readily identifiable: immediate response demands (e.g. parenting a misbehaving child), social norms (e.g. at church), and ethnic culture (e.g. collectivism), respectively. Such distinctions help capture complexity in real situations, where cues from multiple levels are typically combined and point to suitably complex methods (e.g. hierarchical linear modelling) for assessing personality processes.

Fourth, TAT advances links between personality and motivation in two respects. First, traits, as needs (e.g. Murray, 1938), generate intrinsic motivation through expression (i.e. need satisfaction). Second, extrinsic motivation derives from provision of (differentially) valued rewards offered by others (e.g. one's boss). In employment settings, these processes amount to people wanting to work where they are rewarded for being themselves. This basic idea readily generalizes to interpersonal relationships (e.g. husband–wife, friend–friend, and client–therapist), accommodating both supplementary and complementary fit processes serving compatibility.

Fifth, TAT primarily targets personality trait processes but extends to activation of skills and abilities. A key distinction is that skills and abilities are not needs; personality traits, accordingly, are uniquely generative of both intrinsic and extrinsic incentives. Motivation driving application of skills and abilities requires extrinsic reinforcement or companion needs (e.g. achievement striving). Multiactivation processes seem applicable outside of work, where, for example, social skills and cognitive biases might be coactivated with personality traits. TAT thus offers parsimonious integration of assorted situational and personological sources operating in any given situation.

CONCLUSIONS

Promoting Situational Judgement Tests and Assessment Centre Exercises as vehicles for mainstream personality research invites pluralistic advance of interactionist paradigms. At heart is programmatic control of personality-relevant situational characteristics affording plausible generalizability of findings to the real world. Situational specificity of personality-outcome relations observed in the meta-analytic literature underscores the need for careful attention to situational moderators of personality processes. Whether such evidence and key TAT principles bear fruit outside the workplace is largely untested, but if traits and situations operate differently at work than in other settings, it is far from clear what might underlie that distinction.

'Low-Fidelity Simulations' Play Central Roles in Explaining Behaviour

DUSTIN WOOD, GRAHAM H. LOWMAN and P.D. HARMS

University of Alabama, Tuscaloosa, AL, USA
dustin.wood@cba.ua.edu

Abstract: Situational judgment tests (SJTs) are regularly referred to as 'low-fidelity simulations' of situations, which we imagine is intended to acknowledge their limitations. However, the 'low-fidelity simulations' participants engage in when completing SJTs represent forms of mental activity that people perform almost ubiquitously, and such mental simulations can be understood as the proximal causes of a very broad range of behaviours. We briefly illustrate how SJTs can be used in a more elaborate form to make more explicit representations of psychological situations. Copyright © 2017 European Association of Personality Psychology

Much of Lievens's argument for greater use of situational judgment tests (SJTs) within personality psychology focused on their practical and predictive benefits. However, Lievens, like many other researchers, also referred to SJTs on several occasions as 'low-fidelity simulations', which we imagine is intended as an acknowledgment of their limitations—specifically that they 'typically ask people what they would do instead of showing actual behaviour' (Lievens, 2017; Motowidlo, Dunnette, & Carter, 1990).

In this sense, they fall short of the basic goal of better understanding what people *actually* do rather than what they *say* they will do in various situations (Baumeister, Vohs, & Funder, 2007; Furr, 2009).

We do not disagree that SJTs represent 'low-fidelity simulation of situations'. Rather, we argue that these types of simulations represent forms of mental activity that individuals engage in almost ubiquitously when deciding how to respond to situations. That is, when deciding how to

respond to a manager or client, how to ask someone out for a date, what job opportunities to pursue and how to do so, what clothes to wear for the day, or when making almost any other type of decision, individuals are likely to engage in 'low-fidelity simulations of the situation' prior to actually performing any behaviour. These simulations occur when an individual mentally plays out how different choices might transact with the situation and culminate with the individual initiating the action plan they have simulated as producing the best results. These sorts of mental simulations are increasingly understood to be *exactly what the brain has evolved to do* and as types of activity that the brain engages in *all the time* (Clark, 2015; Friston, 2010; Hohwy, 2013). Consequently, we argue that SJTs have underappreciated strengths. In particular, SJTs can be used to provide novel windows into the mental representations of situations, which are widely understood to be the proximal causes of behaviour (Lewin, 1943; Mischel & Shoda, 1995; Reis, 2008).

Lievens's provided considerable detail about how SJTs can be used to provide indirect estimates of an individual's traits, in the form of *implicit trait policies* (see also Motowidlo, Hooper, & Jackson, 2006). However, we imagine there is less awareness that SJTs can also be adapted to provide indirect estimates of a person's *expectancies* (or *beliefs, lay theories,* and *schemas*) about how different actions will interact with situation features, and how these features might transact with one another—key components in understanding mental simulations. These expectancies can be estimated by showing how the individual expects different outcomes to covary across different possible responses to particular situations.

As an example, we asked participants to complete SJTs, which concerned how to respond to a coworker who had done something disrespectful, or contrary to the company's interests (Lowman, Wood, & Harms, 2017). As with standard SJTs, participants then rated how likely they would be to express their disapproval to the coworker. However, participants were also asked how they expected expressions of disapproval to affect various aspects of the situation, such as the likelihood of *getting into a major argument with the coworker*, of *having responded appropriately to the situation*, of *being formally punished* (e.g. getting fired), and so on. In addition, we randomly specified whether the coworker was an equal or lower-ranking employee, or a higher-ranking employee (e.g. a boss or manager). From this type of data, it is possible to construct *functional field* representations of situations, which can be used to provide both visual and mathematically formalized representations of the causal relationships between features within situations, as they may be understood by participants (Wood, Lowman, Harms, & Spain, in press; Wood, Spain, & Harms, in press).

Results indicated that changing the interaction partner within the scenario from 'a coworker' to 'your manager' substantially decreased the likelihood of expressing disapproval. Why? Results indicated that participants generally expected expressions of disapproval to lead to arguments whatever the rank of the coworker in the situation;
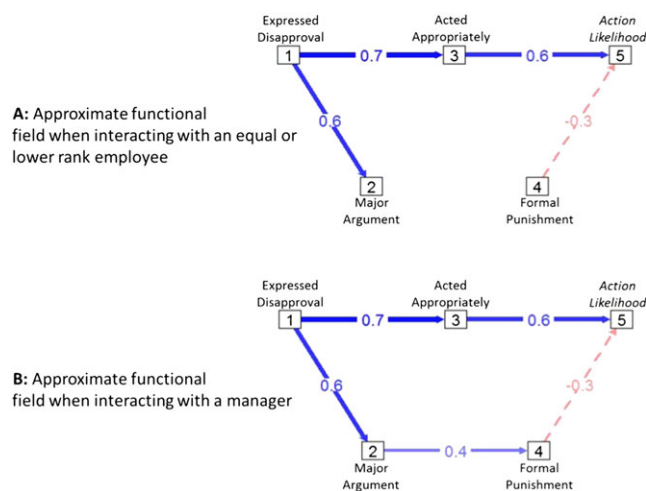

Figure 1. Representation of the functional field linking expressions of disapproval to other outcomes, separated by relative power of the participant and coworker within the interaction. Adapted from results presented in Lowman et al., 2017, figure 6. [Colour figure can be viewed at wileyonlinelibrary.com]

however, these arguments were *only* expected to increase likelihood of getting fired if one was arguing with one's manager. These results are illustrated in Figure 1A and 1B. By using standard 'tracing rules' for path coefficients to estimate indirect effects (Wright, 1934), we can see that the risk of being fired by arguing with a manager produces a negative pathway between expressions of disapproval and the likelihood of performing such actions. This can be understood as meaning that individuals understand expressions of disapproval as being less functional when interacting with a superior than when interacting with a subordinate, specifically by decreasing job security. In this way, the construction of functional field representations from SJT data provides a powerful method for representing the mental simulations individuals undergo to determine how to respond to situations.

An important task for personality and social psychological theory is determining how to represent *psychological situations* better, which is understood to be the proximal cause of people's behaviour (Rauthmann, Sherman, & Funder, 2015; Reis, 2008). The field representations that can be created from elaborated SJTs are sure to prompt many other questions, such as 'how *accurate* are a person's expectancies about how features of the situation relate to one another?' However, in many ways, this kind of 'objective' accuracy is beside the point. Much research on personality disorders and other topics indicates that people's expectancies about situations exert considerable effects on their behaviour irrespective of whether they are accurate (Beck, Freeman, & Davis, 2015). The more important question is whether the field representations that can be generated from elaborated SJTs accurately reflect the 'low-fidelity simulations' that individuals actually perform when deciding how to react to situations. If so, SJTs may be uniquely powerful tools for representing the psychological processes, which underlie behaviour.

Leveraging Situational Judgement Tests to Study Pathological Personality Processes

AIDAN G.C. WRIGHT

University of Pittsburgh, Pittsburgh, PA, USA
aidan@pitt.edu

Abstract: How best to conceptualize, operationalize, and assess personality pathology is the central issue facing personality disorder research today. Lievens's contention that Situational Judgement Tests (SJTs) might have relevance for understanding personality pathology is an interesting suggestion that I believe holds promise. In particular, calculating implicit trait policies from controlled scenarios of relevant SJTs may allow personality pathology researchers to isolate and study specific personality processes. In turn, this may offer new insights into what differentiates normal from abnormal personality functioning. Copyright © 2017 European Association of Personality Psychology

Lievens's article introduced several intriguing ways to apply assessment techniques popular in personnel selection to personality research. One particularly intriguing suggestion was that Situational Judgement Tests (SJTs) might offer new avenues for understanding personality pathology. I think this proposal deserves consideration and offer some elaborations on Lievens's notions, particularly how SJTs might be used to address debates about the very nature of personality pathology.

Descriptions of personality disorders found in diagnostic manuals (e.g. the Diagnostic and Statistical Manual of Mental Disorders, fifth edition; DSM-5) lack validity. That is to say, personality pathology, as manifested in nature, is not organized as 10 discrete categories. Neither does any theory predating the modern DSM era argue that these should be the 10 ways in which personality pathology presents. Instead, review of the empirical literature favours a hierarchical dimensional structure of personality pathology (e.g. Wright & Zimmermann, 2015) that parallels normative trait structure at various levels (Markon et al., 2005; Wright et al., 2012; Wright & Simms, 2014). That structure of personality pathology matches structure of normative traits should be neither surprising nor controversial. Evidence for shared structure answers the question, *how do individual differences in personality disorder manifest*? However, it leaves unanswered, *what differentiates normal from abnormal personality functioning*? Arguably, shared structure makes the second question harder to answer.

Researchers have conceptualized these two questions in various ways. For instance, Pincus (2011) referred to genus and species of personality pathology, Hopwood et al. (2011; see also Sharp et al., 2015) psychometrically isolated shared and unique variance in personality disorders, labelling these components severity and style. In an effort to address these questions and better align the DSM with empirical evidence, the DSM-5 Personality and Personality Disorders Workgroup proposed a revised model that differentiated between what personality disorder is (Criterion A) and individual differences in how it manifests (Criterion B). In the DSM-5 alternative model, which was not formally adopted but was included in the manual in a separate section, Criterion A is instantiated as interpersonal and self impairments, and Criterion B is reflected in a trait model that conforms to a maladaptive hierarchical Big Five structure (Wright et al., 2012; Wright & Simms, 2014).

Although the DSM-5 model has many strengths, it is not without limitations. Most relevant for this discussion is that the trait model is composed of *maladaptive* traits and therefore confounds genus and species. For instance, Criterion A measures rarely increment Criterion B measures when predicting validation variables (e.g. Creswell, Bachrach, Wright, Pinto, & Ansell, 2016; Few et al., 2013), because dysfunction is encoded in the traits already. Furthermore, test information functions of normal range and maladaptive personality trait measures largely overlap for the same latent dimensions (Suzuki, Samuel, Pahlen, & Krueger, 2015; Walton, Roberts, Krueger, Blonigen, & Hicks, 2008). Further complicating matters, personality traits have no privileged association with personality disorders and are in fact strongly related to psychopathology writ large (Kotov et al., 2010). This suggests that what we recognize as personality pathology is not merely extreme manifestations of basic traits but instead reflects (at least partially) alternative processes.

However, establishing workable definitions of generalized personality pathology has proven challenging. I have previously argued that it is not trait levels, per se, that connote personality disorder but rather maladaptive expression (Wright, 2011). To appreciate this fully, one must recognize that traits, whether adaptive range (Fleeson, 2001) or maladaptive (Wright & Simms, 2016) reflect contextualized dynamic processes (e.g. DeYoung, 2015). As such, personality pathology is reflected in not only level but also patterning and match between traits and situations. Notions frequently associated with personality pathology, such as rigidity (i.e. behaving the same across different situations) and instability (i.e. behaving differently across the same situations), are problematic because they contribute to mismatch between an individual's behaviour and what would be effective in situations. More generally, a pattern of behavioural contingencies that sufficiently departs from normative expectations for situations and are dysfunctional in nature might adequately describe personality pathology. Getting at this through standard questionnaire batteries has proven difficult.

As Lievens noted, personality pathology researchers are increasingly turning to ambulatory assessment techniques to study these processes. For instance, in a recent study, several

colleagues and I showed that the more narcissistic individuals were, the more likely they were to respond to perceptions of dominant behaviour in others with negative affect and quarrelsome behaviour (Wright et al., in press). Ambulatory assessment has its own limitations, including lack of situation standardization and heavy participant burden.

SJTs would seem to be ideally suited for studying similar processes, as well as differentiating severity and style in personality functioning. Although this would require considerable effort in development, relevant situations could be designed, along with responses that (a) reflect different trait content (e.g. neuroticism and agreeableness) and (b) different levels of situational effectiveness based on expert consensus. Response patterns of responses would allow identifying trait profiles, as well as levels of maladaptivity specific to each trait domain. Importantly, maladaptivity need not be instantiated in overtly pathological items but rather in matches and mismatches to situational demands. For instance, asserting oneself is important in some scenarios but imprudent in others. In this way, SJTs would allow dissociation of severity and style. They do this by contextualizing variables in situations.

Moreover, variations of implicit trait policies could be used to score overall levels of adaptivity in trait functioning (i.e. effectiveness) and used to predict criterion variables of dysfunction. Separately, overtly problematic items could also be developed. Personality pathology is often ego syntonic, therefore presumably would lead to rating maladaptive range responses as more effective. Finally, as Lievens also suggested, direct tests of concepts such as rigidity and instability in behaviour could be tested. Thus, SJTs and implicit trait theories offer many exciting possibilities for new avenues of inquiry in personality pathology research.

Importance of Testing Validity

MATTHIAS ZIEGLER and KAI T. HORSTMANN

Humboldt-Universität Zu Berlin, Berlin, Germany
zieglema@hu-berlin.de

Abstract: We applaud Lievens for bringing organizational psychologists' work on situational judgment tests and assessment centre exercises to the attention of personality researchers. There are certainly potential fruitful contributions these methods can make. However, we believe that several validity questions need attention first. Copyright © 2017 European Association of Personality Psychology

We applaud Lievens for calling personality researchers' attention to the work on situational judgment tests (SJTs) and assessment centre exercises (ACs) organizational psychologists have conducted over the last decades. The notion of cross-fertilization is consistent with the spirit of Cronbach's (1957) presidential address, even though it was published 60 years ago and referred to experimental and correlational psychology. We agree with Lievens that especially SJTs and ACs are of interest to personality psychology and personality assessment. However, with regard to personality assessment we also see several hurdles that need to be overcome.

Construct validity of the score interpretations of both ACs and SJTs has repeatedly been scrutinized. Empirical results have not always been encouraging and are potential obstacles for their use in personality assessment.

SJTS, COGNITIVE ABILITY, AND THE UTILITY OF THE 'SITUATION'

Krumm, Hüffmeier, and Lievens (2017) suggested using experimental test validation, meaning systematically manipulating fundamental item characteristics to test their roles in response processes. Krumm et al. (2015) also applied this method to SJTs used in a work context and concluded that actual situational description is not necessary for most items. Test takers could conclude the correct answer even without knowing the situation. This phenomenon underscores the overlap between SJTs and cognitive ability. McDaniel, Morgeson, Finnegan, Campion, and Braverman (2001) further reported a meta-analytically derived correlation between SJTs scores and general cognitive ability of .46. This is much higher than the correlations between the Big Five personality domains and general cognitive ability (Ackerman & Heggestad, 1997; all correlations below. 34). Even though cognitive ability was not differentiated into fluid and crystallized intelligence or verbal and spatial in the SJT studies, it seems reasonable to suspect that SJTs are more closely related to crystallized/verbal intelligence. Using indicators of general mental ability might therefore even underestimate the influence of crystallized/verbal intelligence. If this were true, the 2% incremental validity of SJTs (Lievens, 2017) over self-reported personality and cognitive ability would need to be replicated when also controlling crystallized/verbal intelligence. Otherwise, it would seem questionable to replace self-reports with SJTs.

WHAT ACS MEASURE

The problematic construct validity of AC scores is anything but new in the literature (Sackett & Dreher, 1982). Kuncel and Sackett (2014) recently showed that AC scores are strongly influenced by a general factor. They also interpreted this general factor as likely related to cognitive ability or conscientiousness. Thus, as is the case

for SJTs, the construct validity of AC scores, especially distinctiveness from cognitive ability needs to be improved. Of course, we recognize that many of the studies indicating problems did not evaluate SJTs or ACs measuring personality, so the applicability of their findings to these might be limited.

TEST-CRITERION CORRELATIONS AND LEVEL OF SYMMETRY

Brunswik (1955) noted the role of symmetry between variables in their correlations. This applies to interpretation of test-criterion correlations (Ziegler & Brunner, 2016). Test-criterion correlations can be smaller when one variable is more abstract than the other. SJT (especially with small numbers of items) and AC scores are probably often less abstract than personality measures. After all, behaviour is aggregated across a small number of (very specific) situations. Trait self-report items usually are much more abstract and so are the resulting scale scores. While a less abstract score per se can be advantageous for test-criterion correlations (Mõttus, 2016), differences in the level of abstraction potentially distort comparisons among different predictors. For example, Meriac, Hoffman, Woehr, and Fleisher (2008) reported incremental validities of AC scores above cognitive ability (see above) and Big Five scores. Considering the differences in symmetry between AC scores and Big Five scores, a fairer comparison would have been to use Big Five facet scores, and similar arguments could be made for SJT scores.

ACs and SJTs can be useful tools in the personality psychologist's toolbox. We generally encourage embracing the possibilities presented by ACs and SJTs, yet the validity concerns regarding their test-scores call for thorough examination first and for each of the intended purposes separately.

AUTHOR'S RESPONSE

Integrating Situational Judgment Tests and Assessment Centre Exercises into Personality Research: Challenges and Further Opportunities

FILIP LIEVENS
Ghent University, Ghent, Belgium
filip.lievens@ugent.be

Abstract: In this rejoinder, I discuss several broad themes that emerged from the 32 commentaries in response to my target article about the value of integrating Situational Judgment Tests and Assessment Centre exercises for advancing research on the personality–situation interplay. Specifically, I summarize the various challenges put forward (e.g. deficient/contaminated trait measurement, assessment of situation construal, inclusion of personality-driven situation experience) and offer potential conceptual (e.g. use of situational taxonomies and contextualized personality perception) and methodological solutions (e.g. use of open-ended and nonlinear SJTs and multiple speed assessment). Taken together, this fits into a more comprehensive multi-method approach to personality assessment. Copyright © 2017 European Association of Personality Psychology

When I was preparing the first version of my article as a keynote for the European Conference on Personality in Lausanne, some of my industrial and organizational psychologist colleagues suggested that I was entering a lion's den. After all, I was speaking (as an industrial and organizational psychologist) about personality assessment to an audience of personality researchers. But being open to new experiences ('Every day do something that scares you'), I was looking forward to this opportunity. I also tremendously enjoyed the constructive and enriching interactions after the keynote. When the comments on the written version of my article trickled in, the same positive vibes returned. I am therefore very grateful to all authors who wrote such thoughtful, insightful, and open-minded comments. I agree with many of the ideas expressed in them and also discovered many useful new ones.

The purpose of my article was to set up a constructive dialogue about the viability of incorporating Situational Judgment Tests (SJTs) and Assessment Centre exercises (ACs) in personality research. I concentrated on these two procedures because they have yet to be taken up by personality researchers and because they match well with a focus on personality–situation interplay. In particular, I aimed to show how these assessment approaches permit tackling key issues such as within-person variability across situations, trait-behaviour linkages, and trait expression and perception.

The dialogue got off to an excellent start with the 32 comments. They also provided a first important test of the viability of integrating the two approaches in personality research. Throughout, there was general agreement on many points. First, scholars mentioned that personality psychology has

remained too insulated from industrial and organizational psychology and that cross-fertilization between these domains has considerable potential to extend knowledge on the aforementioned topics. Second, need for more diversity in the existing methodological approaches for assessing personality was acknowledged. Third, the emphasis on ecological validity and on behaviour in real-world settings in personnel selection was applauded. Finally, the two assessment procedures were regarded as providing unique opportunities for better understanding the personality triad (persons, situations, and behaviours). At the same time, the commenters also noted challenges to be met, barriers to be overcome, and additional opportunities for research and collaboration. In this rejoinder, I am not able to address all specific issues mentioned throughout the individual comments. Instead, I focus on several broad themes that emerged from them.

SHOULD ONE PUT EFFORT INTO CONTEXTUALIZING PERSONALITY ASSESSMENT?

A common element in SJTs and ACs is that they assess traits in more contextualized formats. Selection researchers started adopting these methods (beyond self-reports) because they have potential to lead to better predictions, to be less susceptible to faking good, and to receive more favourable applicant perceptions. Some commenters supported this contextualized approach with additional arguments. For example, Durbin and Hicks applauded the contextualized approach inherent in SJT/ACs because it is both deep (indicative of hypothetical response choices in SJTs and of behaviour in ACs) and narrow (responses relate to particular life situations). In addition, Baumert and Blum made a case that reliance on contextualized items from a variety of contexts is more likely to capture a construct in its full range than self-reports of global traits because the latter require people to average their behavioural tendencies mentally across many situations. Hopwood and Bleidorn mentioned that contextualized measures and contextually relevant traits might be more useful to predict personality changes than applying broad measures multiple times.

Other commenters raised critical voices about the need for contextualization. Mõttus argued it is premature to invest in newer more contextualized measurement methods. Instead, he suggested looking for the best possible uses of already existing methods to improve predictive accuracy. Specifically, he pleaded for development and validation of item-based predictive models via machine learning (instead of using a priori composite trait scores). Although this view deviates from how personality has been measured for decades and might open the door to dust-bowl empiricism, I agree that incorporating machine learning in personality measurement is one of the main opportunities in the years to come. In addition, I stress that machine-learning approaches do not preclude using contextualized personality items (alongside generalized ones).

Christiansen and Speer cautioned that increased contextualization in SJTs and ACs might lead to lower validity because the context decreases bandwidth (reflecting a narrower range of demands) and/or the judges are less accurate. Funder argued that the goals of personality psychology and applied domains such as industrial and organizational psychology (personnel selection) are not the same: personality psychology focuses more on understanding than prediction and therefore benefits from using global traits that go beyond specific contexts, whereas the opposite is true for more applied domains. Funder argued therefore that too much contextualization might jeopardize the explanatory value of personality; a point that was also made by Leikas.

I believe Funder's distinction between prediction and explanation should be given full attention. Yet I am hesitant to equate personality and industrial/organizational psychology with the two extremes of his continuum because neither discipline maintains (e.g. social effects of personality in the personality domain) a pure focus on either explanation or prediction. In addition, research demonstrates that contextualizing personality inventories does not necessarily lead to poorer prediction of broad outcomes. On the contrary, meta-analytic research revealed that contextualized personality self-reports predicted overall job performance twice as well than did global non-contextualized self-reports (Schaffer & Postlethwaite, 2012). In any case, I echo the calls of many commenters for more research on contextualized personality measures in various contexts.

SJTS IN PERSONALITY RESEARCH: CHALLENGES AND OPPORTUNITIES

My article put a heavy emphasis on SJTs as new innovative approaches for advancing personality research on within-person variability, trait-behaviour linkages, and personality disorders. Overall, the commenters concurred that the SJT format is a cost-effective (albeit ecologically less valid) alternative to experience-sampling, thereby allowing more control in examining within-person variability and situation-trait contingencies via the standardization of situations. As noted by Corr, an integrative picture emerges by integrating situationally driven intra-individual differences (variability) with inter-individual differences (diversity), which is consistent with Whole Trait Theory (Fleeson & Jayawickreme, 2015). Commenters also saw potential in using situational taxonomies to build SJT scenarios and examining individual differences in situation construal. This might provide insight into people's mental models and how they relate to (mal)adaptive behaviour. Yet the commenters also pointed out several challenges and opportunities for future research.

TRAIT MEASUREMENT IN SJTS

Some scholars expressed concerns about SJT scores' construct-related validity. Essentially, these concerns dealt with deficiency and contamination in them (Judge, Hofmans,

& Wille; Ziegler & Horstmann). Ziegler and Horstmann referred to potential contamination of SJT socres by crystalized intelligence. In a similar vein, Judge et al. warned that many SJTs are heterogeneous at the item level so that their scores typically correlate with a variety of constructs that are often not intended to be assessed, thereby making the scores less than clear measures of any personality construct.

I agree that construct measurement in SJTs should not be taken for granted. There are at least two aspects that one should note. First, at a conceptual level, it is pivotal to keep the underlying theory behind SJTs in mind. As delineated in the theory of knowledge determinants of SJTs (Lievens & Motowidlo, 2016; Motowidlo & Beier, 2010; Motowidlo, Hooper, & Jackson, 2006), SJT scores reflect procedural knowledge (in the form of implicit trait policies). So, in line with Motowidlo's comment, I stress again that implicit trait policies are not equivalent to personality traits because a person's implicit trait policy represents his or her procedural knowledge about relations between expressions of traits and their effectiveness in situations to achieve some goal. Motowidlo further clarified that the trait (e.g. agreeableness) is one of the antecedents of the implicit trait policy for a given trait.

Second, there is the measurement aspect. Although I agree that construct measurement was an Achilles Heel of many early SJTs, recent research with construct-driven SJTs provides evidence that it is possible to assess specific well-defined traits (i.e. people's procedural knowledge about the effectiveness of these traits; e.g. Bledow & Frese, 2009; Motowidlo et al., 2006; Mussel, Gatzka, & Hewig, 2016). In particular, I refer to Mussel et al., 2016. for an excellent example. They constructed an SJT with acceptable convergent and discriminant validities between self and other reports of analogous personality traits. Key 'ingredients of their recipe' to design such an SJT included reliance on clearly defined personality traits instead of vague dimensions and development of response options to reflect low versus high trait levels (see Lievens, in press for a more general overview of construct-driven SJTs).

SITUATION CONSTRUAL AND SJTS

Several commenters shared my enthusiasm for using SJTs (and especially implicit trait policies) as novel approaches for assessing situation construal because situation construal is generally recognized to be a proximal cause of behaviour (Funder, 2016). There was agreement that measuring it via SJTs might improve explanation and understanding of people's behaviour, cognitions, and feelings. For instance, according to Wood, Lowman, and Harms, individuals engage every day in a large variety of 'low-fidelity situation simulations' (e.g. when choosing what clothes to wear) prior to engaging in them, which underscores the relevance of SJTs. Baumert and Blum referred to reactive transactions that can be assessed by varying situational features in SJT scenarios and by examining how people differ in how they perceive the same scenarios (and act upon them when an SJT item's situation occurs in real live). I echo that assessment of

situation construal is of pivotal importance in promoting SJT use in the personality domain due to its role in many theories (e.g. CAPS).

That said, some caveats and challenges related to the use of SJTs and situation construal were also mentioned. Motowidlo qualified my suggestion that SJT items can be developed for wide arrays of life domains (e.g. school, health, romantic relationships, and sports). He posited that SJTs' notion of implicit trait policies would be most useful for predicting goal-oriented behavioural patterns that are intrinsically functional or dysfunctional for individuals or society.

Wood et al. argued that implicit trait policies and situation construal as captured via SJTs do not represent psychological situations in their full ranges. They proposed the notion of field representations to capture psychological situations more comprehensively. People's functional field representations of the situation were defined as the mental simulations and expectancies about how situation features covary to determine how to respond to the situation. To assess functional field representations via SJTs, Wood et al. proposed prompting people not only about actions they would choose but also about their evaluations of how their chosen actions would affect specific situation aspects. Similarly, Durbin and Hicks stressed that SJT measurement should be widened by incorporating people's schemas/expectations regarding the likely outcomes of selected responses.

To improve assessing respondents' situation construals further, Rockstuhl made a strong plea for development and use of open-ended SJTs. Such SJTs do not present lists of predetermined response options to individuals. Instead, they ask people for free responses. This can be done in writing (written constructed responses), orally (via think-aloud, see Durbin & Hicks), or via patterned behaviour description interviews (see Heimann & Ingold). Moreover, Rockstuhl posited that situational construal can best be assessed when SJTs manipulate specific situational cues on the basis of situational taxonomies and that such an approach might expand understanding of situation–trait contingencies.

I agree that extensions of the typical SJT format such as open-ended SJTs (especially if they are coupled with manipulations of situational features) are useful to focus on how people construe the standardized scenarios, their memories of similar situations, and their rationales for different responses. At the same time, I note that multiple-choice format does not per se preclude assessing situation construal. It is also possible to measure situation construal via multiple-choice format (e.g. How do you judge the situation? Choose one of the four responses below).

PERSONALITY-DRIVEN SITUATION EXPERIENCE AND SJTS

Several scholars (Lezotte, Condon, & Mroczek; Lilienfeld; Wood, Lowman, & Harms) mentioned the 'contrived' nature of SJTs as a potential limitation. In particular, a common thread running through many commentaries (Baumert & Blum; Judge, Hofmans, & Wille; Lilienfeld; Oh, Kim, &

Kim; Rauthmann) was that SJTs typically present people with pre-determined sets of SJT items. Thus, all people receive the same standardized situations in the same sequence, which does not match how actual situations unfold. I acknowledged this limitation inherent in the traditional make-up of SJTs in my article. Thus, I concur that one of the greatest advantages of SJTs, their standardized stimulus and response format, might also be their biggest liability. I also agree that the prototypical SJT mainly gets into personality reactivity (Judge et al.) and suppresses variance in personality-driven situation experience (Rauthmann, Sherman, Nave, et al., 2015) because it constrains individuals' propensities to select situations, shape existing situations, and create novel situations. As discussed by Judge et al., the standardization imposed by SJTs might also present situations to people that are not representative of those they typically encounter. According to Oh et al., this drawback is especially important for assessing personality traits such as proactive personality (but see the personal initiative SJT developed by Bledow & Frese, 2009).

I propose at least two approaches to deal with this limitation. First, I suggest adopting nonlinear SJTs in which respondents' answers to earlier items influence the subsequent situation items they receive. Baumert and Blum also proposed this. Conceptually, they argued that non-linear/branched/game-like SJTs might get beyond reactive transactions and dig deeper into proactive/manipulative/evocative transactions because they would offer opportunities to enter, avoid, or shape specific situations. As a potential downside, this non-linear and gamified approach challenges adequate construct measurement in SJTs even more. I concur with Handler (2013) that this trade-off between realism and trait measurement will be one of the key challenges in the years to come. Or as Handler aptly put it: 'The use of simulations forces us to choose between raw empiricism that does not provide sound trait-based measurement and highly structured and less fluid simulations, that while measuring important traits, place limitations on realism and complexity. I believe that the future lies in bridging this gap.' (p. viii).

Second, as SJTs are low-fidelity simulations of what people say they will do in different situations (instead of what they actually do), Breil, Geukes, and Back suggested using SJTs and ACs in tandem as a promising means to understand intra-individual variability better in controlled settings. They proposed comparing discrepancies between people's procedural knowledge (assessed via SJT) and their actually expressed behaviour (assessed via AC exercises or webcam SJT).

OTHER PERSONALITY LEVELS AND SJTS

According to Dunlop and Horton, SJTs are based on the narrow assumption that personality primarily consists of trait-based characteristics. They argued that the SJT paradigm should be extended to include other personality aspects. I agree that SJTs focus on personality traits. However, there exist various options to widen the SJT paradigm to include other aspects such as motivations/goals

and narratives. Regarding narratives, Dunlop and Horton proposed that SJTs should also identify the stories people associate with the scenarios. This is indeed possible. In fact, in personnel selection, such elaboration has been used to reduce faking good on SJTs (Lievens & Peeters, 2008). Regarding motivations/goals, Dunlop and Horton suggested optimizing SJTs by also identifying the motivations/goals people have for the scenarios. Drawing on the notions of multi-finality (i.e. a behaviour usually has more than one potential consequence) and equi-finality (i.e. different behaviours often have identical consequences), Pretsch and Schmitt made similar suggestions to extend the SJT format by explicitly crossing several traits with several goals in SJT items. Breil et al. proposed designing SJT response options so that they represent different strategies in goal achievement. I am grateful for these suggested extensions because they not only permit obtaining insight into the specific goals that motivate people to choose different trait-related responses but also allow studying rigidity or flexibility in people's trait expressions across goals and situations.

Some commenters (Christiansen & Speer; Heimann & Ingold) went one step further and suggested instruments other than SJTs that might be even better suited to tap into people's goals/motivations and narratives. Christiansen and Speer posited to use biographical questionnaires as alternative personality measures. When these biographical inventories are open-ended (e.g. in the form of essays), I agree that they might further widen the spectrum of personality instruments and tap into personality levels other than traits (i.e. narratives). Yet to obtain insights into the personality–situation interplay, it is crucial that these biographical questionnaires be contextualized to include situational demands that correspond to the demands that affect the behaviours one wants to explain/predict.

Heimann and Ingold made a compelling case for including patterned behaviour description interviews in the personality-research toolbox. In these interviews, people are asked about their trait-related behaviours in situations that they experienced in the past. Unlike SJTs, patterned behaviour description interviews apply an open-ended format wherein participants present situations they encountered and interviewers use prompts to get clear pictures of how participants behaved in them. Patterned behaviour description interviews have several advantages. Apart from self-descriptions of behaviour, the short narrative presented in the interviews provides also information on people's cognitions, emotions, and motivations for choosing specific behaviours and the goals they aim to achieve. People's situation selections also tell something about their personalities. Finally, these interviews allow assessing actual communication behaviour.

Given that such interviews might provide information on the three levels of personality mentioned by Dunlop and Horton, I agree that these interviews further diversify personality assessment methods. That said, I also note that the benefits of patterned behaviour description interviews come with a price of lower stimulus presentation (each interviewee presents different situations and interviewers vary

in prompt level and content) and response scoring consistency (interviewers rate people's answers). I am also less enthusiastic about the viability of patterned behaviour description interviews for studying within-person variability because such interviews typically tap fewer situations than do SJT items.

PERSONALITY DISORDERS AND SJTS

Generally, my suggestions to use SJTs for advancing measurement and understanding of personality disorders were well received. Several commenters (De Fruyt, De Clercq, Verbeke, & Vergauwe; Durbin & Hicks; Wood et al.; Wright) concurred with the various advantages that I mentioned when using SJTs in this domain as additions or even alternatives to structured clinical interviews, self-report questionnaires, and ambulatory assessment approaches. SJTs' advantages include elicitation of individual differences in how patients interpret and respond to sets of standardized scenarios, use of situational taxonomies (e.g. interpersonal circumplex and DIAMONDS) for building these scenarios, ease of administration, and automated scoring. I echo the commenters that SJTs can reveal relevant diagnostic information (and eventually assess personality disorders) by detecting someone's (i) endorsement of overtly problematic response options, (ii) situation–trait contingencies that mismatch those of 'normal' people, and (iii) discrepancies between responses to 'what should you do' and 'what would you do' instructions. Future research should compare the convergence among these approaches in detecting personality dysfunctions.

It was encouraging that De Fruyt et al. have already developed an SJT for assessing borderline personality pathology. This confirms the feasibility of the SJT approach in this area. Durbin & Hicks suggested adding to SJTs think-aloud procedures (for digging into situation construal) and/or behavioural observation (for examining whether people with more maladaptive personality structure have greater discrepancies between knowledge of effective trait-related behaviour and actual expressions of it). Relatedly, Lilienfeld posited that the standardized nature of AC exercises might avoid the lower ecological validity of SJTs and test even better the notion that many personality disorders are inherently interpersonal. Lilienfeld made an especially compelling case for using multiple AC exercises to ensure that the positive impressions that for example narcissists tend to create in the short run (Back, Schmukle, & Egloff, 2010) do not deceive observers.

Finally, at a conceptual level, Wright made a point that I should have highlighted more in my article. He argued that the viability of SJTs in this area is contingent upon a workable definition of personality pathology. That is, SJTs are especially useful if one conceives personality dysfunctions as mismatches between people's trait expressions interact with situational demands (instead of extreme trait levels per se). I concur with Wright that such a definition determines how personality disorders manifest themselves SJT responses (e.g. rigidity across different situations, instability across similar situations, or situation–trait contingencies that deviate from normative/adaptive patterns).

Importance of relying on an underlying theory of SJT development is also evident in Jayawickreme's proposal for using SJTs to identify individuals who are more likely to experience positive changes after traumatic experiences. For SJTs to predict this, theories about post-traumatic growth should guide their development. Theorizing in this domain informs which adverse situations to sample, which response options to include, and how to score participants' responses (e.g. people that systematically choose options in which they seek social support receive higher scores). When experience-sampling studies serve as criterion measures, an additional benefit of such theory-driven SJTs is that they allow testing specific theoretical assumptions underlying post-traumatic growth.

QUANTIFICATION OF WITHIN-PERSON VARIABILITY IN SJTS

Although my idea for using SJTs for capturing within-person variability was favourably received, questions were raised about how to quantify and model this variability (Breil et al.; Corr; Lang, Tackett, & Zettler). Along these lines, Lang et al. provided excellent analytical recommendations. They made a case that using the traditional standard deviation (SD) as a within-person variability index in SJTs might confound various sources of variance: the SD might not only reflect degree of variability with which a person shows a tendency to choose responses related to a given trait across situations; it might also be affected by the social presses of the situations (items), the respondee's trait levels, and individual differences in careless responding.[4]

To avoid these confounds Lang et al. argued convincingly in favour of using Item Response Theory (i.e. tree models) for modelling within-person variability. The key advantage of this IRT-based approach is that it splits information from SJT responses into (i) persons' mean-trait levels/latent traits, (ii) persons' tendencies to express the trait variably, (iii) items' mean-level/latent difficulty, and (iv) items' tendencies to elicit rating variability. As a key implication, this IRT-based approach (as opposed to the classic SD approach) permits conceptual clarity because it disentangles people's intra-individual variability from other sources of variance. I thus fully concur that advanced psychometric models represent important and complementary building blocks for research on within-person variability via SJTs. Therefore, they should also be integrated in the Breil et al. framework that distinguishes between within-context variability and cross-context variability.

[4]Compare this to the SD in experience-sampling studies that use the same scales each day. As these scales have larger numbers of items/observations, ratings are less likely affected by situation/item press confounds.

ACS IN PERSONALITY RESEARCH: CHALLENGES AND OPPORTUNITIES

My article made a case that AC exercises might advance research on personality expression, personality perception, and their transactions. Generally, commenters agreed with the areas of research that I identified but also raised some concerns and suggested additional research ideas, which I elaborate on below.

ACS AS PERSONALITY MEASURES

Similar to SJTs, several authors (Johnson; Judge et al.; Leikas; Pretsch & Schmitt) questioned whether AC exercises assess personality. According to Johnson and Leikas, for example, the behaviour displayed by candidates in AC exercises is not reflective of their standings on personality traits. Instead, the behaviours shown reflect their skills in displaying the behaviour that they consider important for the situations (whether specifically what the AC judges are looking for, or what would be appropriate to the situation and they would do in real life). Judge et al. suggested that the behavioural reactions evoked in these assessments often reflect constructs that do not belong to the personality domain, thereby creating construct contamination.

Relatedly, various authors (Heimann & Ingold; Judge et al.; Mussel, et al.) also raised questions about the construct-related validity of AC ratings. They pointed to the large research base showing that ratings from given AC dimensions do not correlate with ratings of the same dimensions across exercises (e.g. Bowler & Woehr, 2006; Lievens & Conway, 2001). I do not see this robust finding as a limitation. The limited convergence of people's ratings on the same dimension across different ACs is consistent with research on levels of cross-situational consistency across situations when these situations have differing demands (Christiansen, Hoffman, Lievens, & Speer, 2013). In fact, this finding shows that ACs are ideally suited for examining how the interplay of personality and situation shapes people's behaviour.

TRAIT ACTIVATION AND ACS

In my article, I discussed how Trait Activation Theory has been fruitfully used to identify which AC exercise features trigger candidate behaviour. Tett offered additional arguments for extending Trait Activation Theory to personality research outside the workplace. As the key principles of trait activation theory outside the workplace have so far been largely untested, he suggested identifying the situational cues (at the task, interpersonal, and contextual levels) that evoke trait-related behaviour in a wide set of life situations (e.g. expression of road rage, prosocial behaviour). Consistent with Rauthman, I also suggest relying on recently developed situational taxonomies for structuring the variety of situational cues. Such taxonomies provide ample opportunities to vary

situation cues to evoke differences in trait-related behaviours (Breil et al.). Rauthmann's situation strength matrix is also an excellent example of how personality researchers might vary situations to detect individual differences.

Some commenters, however, tempered enthusiasm for the trait activation logic in AC-like tasks. For instance, Pretsch and Schmitt pointed to potential observability differences among traits (aka hierarchy of trait visibility), thereby questioning whether several simultaneously activated traits are equally visible to observers. They also noted that, especially in group discussions, individuals' trait expressions might depend on other people's trait expressions. I acknowledge this caveat. ACs are diverse and range from single-participant exercises, to one-on-one exercises, to group discussions. The constructs targeted and assessed vary across these formats. When traits are less observable in given exercises, recent research has demonstrated that planting specific stimuli to evoke desired trait behaviours in the exercises (e.g. via role-players or by changing contextual features) and familiarizing assessors with them increases assessor detection and utilization of these behaviours in scoring (Lievens et al., 2015). Breil et al. concurred that role-players and other interaction partners might produce the needed varying and adaptive situational cues (in addition to the fixed ones that are already part of initially presented situations).

IN SITU, EX SITU, AND JUXTA SITUM ASSESSMENT

In my article, I did not elaborate on whether raters are present when candidates participate in ACs and/or whether afterwards they watch videotaped candidate performances. Whatever option is chosen in personnel selection, personality researchers could use any combination of various rating approaches. Along these lines, Rauthmann made useful distinctions among three (not mutually exclusive) options: in situ, ex situ, and juxta situm raters. In situ rating refers to participants rating their own performances, whereas ex situ raters are a posteriori raters/coders who evaluate videos of the participants. In current AC practice, combinations of these ratings are common. In online ACs, ex situ raters are often used. On-site assessors who are physically present when the participants participate in the ACs can be regarded as juxta situm raters. In some cases, fellow AC participants and role players might also serve as AC raters.

In prior personality research (e.g. Rauthmann, Sherman, Nave, et al., 2015), juxta situm raters were often not available. However, when ACs are used, it is possible (and often recommended) to have combinations of in situ raters, ex situ raters, and juxta situm raters. This triangulation of raters is of pivotal conceptual importance: it enables researchers to disentangle variance due to objective situations, situation construal, and people's traits. In turn, partitioning the variance according to these different factors allows more comprehensive tests of situation construal models (e.g. Funder, 2016).

PERSONALITY-DRIVEN SITUATION EXPERIENCE AND ACS

Contrary to SJTs, AC exercises give people discretion to change and shape situations during the exercises. Unfortunately, current AC practices do not fully capitalize on this advantage because assessors typically rate people's behaviour but not their strategies to shape the situations. Therefore, I echo Rauthmann that it is vital to evaluate the extents to which people change, evoke, or create new situations in AC settings (see also Rauthmann & Sherman, 2016c).

In particular, as discussed in some comments (Gibbons & Rupp; Leikas; Rauthmann), the interpersonal and dynamic features of ACs provide unique opportunities for examining how people shape interpersonal situations and create new ones. Therefore, I concur that ACs (e.g. group discussions and role plays) are ideally suited for testing principles underlying interpersonal theory by systematically varying situational characteristics such as the numbers, status, sexes, intentions, personalities, and so forth of the other participants involved. Thus, the social and dynamic nature of ACs should be viewed as real opportunities. Compared to other instruments (e.g. SJTs), the ecological validity for making inferences related to interpersonal competencies is substantial in ACs, although some losses in internal validity should be acknowledged (Judge et al.; Pretsch & Schmitt).

MULTIPLE SPEED ASSESSMENTS

Some commenters (Lezotte et al.; Lilienfeld; Wood et al.) noted that behavioural variability unfolds over time (over days and weeks) and wondered whether and how this temporal dimension can be included in assessment procedures other than experience-sampling. In my article, I gave the example of how a multiple speed assessment procedure (18 short 3-minute role plays) was set up to assess people's interpersonal competencies (Herde & Lievens, 2016). Per role play, there was one role player who also served as assessor. Such multiple speed assessment enables assessing how people vary their behaviour. Variation can then be conceptualized as behavioural flexibility across the various exercises and/or as 'learning' throughout the session (Baard, Rench, & Kozlowski, 2014).

De Fruyt et al. elaborated on this multiple speed assessment idea. They pointed out how large sets of brief AC-like exercises might be fruitfully used to evaluate personality change after specific interventions (e.g. coaching on expressing specific traits). As an advantage, in this assessment procedure, personality change can be examined independently from the people who received the interventions by making each exercise's assessor a 'blank slate' (i.e. unfamiliar with the participants and blind to the coached traits).

Finally, Fleeson and Hamza built on the multiple speed assessment idea and on the notion of successive winnowing to propose a 'Successive Situation-Based Selection' approach. This intriguing approach permits personality researchers to capitalize on the advantages of behavioural

assessment while at the same time reducing some of the typical costs. That is, Fleeson and Hamza suggested gradually reducing initially undifferentiated samples of participants to more extreme (high and low) groups via successive selections on the basis of performance in various ACs. Next, these extreme groups allow researchers examining dimensions of interest (e.g. integrity and dark triad) with more power and less assessment time. Although this procedure does not shed light on variability across time, I agree that it represents another example of how the AC methodology can be flexibly used. I also anticipate that in the future, social-sensing techniques (Schmidt Mast, Gatica-Perez, Frauendorfer, Nguyen, & Choudhury, 2015) and virtual (remote/online/gamified) ACs might even replace role players/assessors to save costs (see also Lezotte et al.).

CONTEXTUALIZED PERSONALITY PERCEPTION

In my article, I laid out the many parallels between AC research and the zero-acquaintance literature. I argued in favour of better cross-fertilization between these two literatures that have evolved almost independently. I discussed among other things how ACs and novel developments related to good judgement (dispositional reasoning and transactions between good judges and good targets/good information) might be especially beneficial for personality research.

Generally, commenters (e.g. De Kock; Funder; Gibbons & Rupp; Letzring & Colman; Mignault & Human) reacted enthusiastically to the prospect of embedding the systematic AC into behavioural observation research in the personality domain. Letzring and Colman discussed that such research is needed to test whether results obtained in zero-acquaintance research still stand in real-world situations where the stakes and motivations of both target and judge are higher. According to Corr, the combined attention to personality expression and perception underscores the notion that traits are socially contextualized. Thus, in addition to people's personality expression across situations, how these expressions are judged in ACs provides unique information about targets.

The commenters also built on my suggestions and offered other opportunities for joint research. The most striking common thread was that they suggested adopting more contextualized approaches to personality perception. For instance, De Kock suggested moving from examination of what makes people good judges of others to what makes people good judges of people *in their situations*. He thus argued that contextualized approaches to personality assessment deal not only with focal persons but also with how people judge persons in situations and in person–situation transactions. Only when this part of the assessment process is taken into account will deeper understanding of personality perception be obtained. To this end, De Kock called for developing a measure of situational reasoning (besides the already existing one on dispositional reasoning) and provided various ways researchers might accomplish this. This situational reasoning measure might extend the current
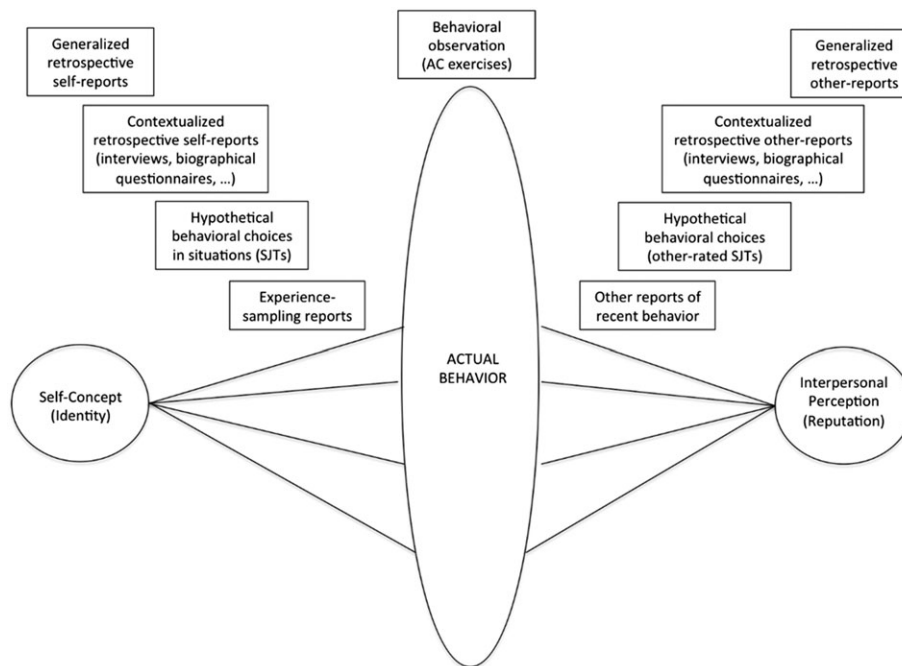
Figure 1.   A lens model representation of multiple methods to assess personality (see also Back & Egloff, 2009).

contextualization measure that is included in dispositional reasoning (De Kock, Lievens, & Born, 2015).

Gibbons and Rupp also stressed the need for contextualized approaches to personality perception among assessors and in assessor training programs. I agree that current training programs are based on traditional trait psychology and seldom take person–situation transactions into account. Assessors should thus be taught to be more aware of contexts in which behaviours occur. Only then they will be more appreciative of people's *if ... then ...* signatures.

Apart from the advantages mentioned by Gibbons and Rupp, a contextualized approach to personality perception could also inform research on other reports (e.g. from coworkers, family members, and friends). Some scholars (Christiansen & Speer; Oh et al.) noted that such other reports deserved more attention in my article. Although other-reports of target people's personalities, as measures of reputation, have incremental validity over self-reports in predicting important outcomes, this aspect is typically not evaluated as part of the interplay between personality and situation. Interestingly, Christiansen and Speer discussed recent research (Kluemper, McLarty, & Bing, 2017) that attested to the validity benefits of using observer ratings from relevant contexts. We need more studies with other reports that take contextualization into account.

To examine accuracy of such contextualized personality judgements, Letzring and Colman proposed an innovative approach. Contrary to the common practice of measuring person perception accuracy as agreement between target people's standings on given traits and judges' assessments of them, they suggested asking judges to complete SJT items, which ask which course of action target people would choose to respond to the situations. Target people's/acquaintances'

responses to the SJT items serve as criteria to determine how accurately judges perceive targets' identities/reputations. I welcome this new approach because (i) number of correctly identified SJT responses represents an objective accuracy benchmark, (ii) it pays more attention to the situations in which personality judgments are made, and (iii) it increases the range of methodologies used in accuracy research.

Finally, Mignault and Human made some suggestions for how research evidence on the good target and expressive accuracy might improve judgement accuracy in personnel selection. I agree that this is largely unchartered territory for selection researchers, who have typically focused on good judge factors. Mignault and Human discussed that it is easier for assessors to rate candidates who can better showcase their qualities. Hence, one recommendation was that assessors should create warm and friendly atmospheres. To improve accuracy Mignault and Human also suggested fostering individuals' self-presentations because research shows it improves people's clarity of expression and judges' cue detection. This runs somewhat counter to common assumptions of selection researchers that self-presentation is often equated with artificial impression management and socially desirable responding. Future studies should pit these competing rationales against each other.

SJTS AND ACS IN MULTI-METHOD APPROACHES TO PERSONALITY ASSESSMENT

As many commenters noted (e.g. Connelly & McAbee; Corr; Heimann & Ingold), integrating SJTs and ACs into personality research fits into comprehensive multi-method approaches to personality assessment that capitalize on

advantages of all methods included (such as traditionally reliable and cost-effective self-reports[5]).

I endorse the multi-method approach to personality assessment. Yet applying it means that several conceptual, methodological, and statistical questions must be addressed. For starters, it begs the questions what one should do with the traditional measures and, for instance, how one should integrate SJTs and ACs with these measures. I echo Connelly and McAbee's useful distinction between using new measures such as SJTs or ACs as (i) replacements (i.e. using new measures instead of old measures), (ii) supplements (e.g. summing scores across measures to produce more precise and representative multi-method measures), or (iii) complements (i.e. using new measures alongside old measures).

I support the complementary approach because there should not be competition between old and new methods and all measures validly tap varying aspects of similar constructs (see also Durbin & Hicks; Hopwood & Bleidorn). Thus, new assessment methods provide information that partially overlaps with that of traditional ones but also allow assessing previously untapped and unique elements of personality. To this end, Connelly and McAbee outlined a useful structural equation modelling approach to tease out variance shared among measures from variance unique to individual measures.

Conceptually, the complementary perspective is consistent with a broad personality conceptualization that does not equate personality only with traits but also with goals/motivations and narratives (Dunlop & Horton). The complementary perspective is also consistent with a lens model that distinguishes between personality measurements related to (i) behavioural observation, (ii) self-concept, and (iii) interpersonal perception (Back & Egloff, 2009). It is therefore worthwhile to update Back and Egloff's model with the new measures discussed in my article and in the comments (SJTs, ACs, but also patterned behaviour description interviews, biographical questionnaires). Figure 1 presents this updated model.

EPILOGUE

In my article, I showcased two personnel selection procedures (SJTs and ACs) that have largely escaped personality researchers' attention. The subsequent 32 commentaries as well as this rejoinder attest to the start of a constructive dialogue between selection and personality researchers that should enrich the theoretical and empirical approaches of both areas. This fits well into the broad idea that psychology researchers should leave their silos and cross the borders of their own respective disciplines. Once again, I would like to thank all the commenters. Key initial steps have been taken to formulate and refine a joint research agenda. I am therefore confident that both SJTs and ACs will inspire personality researchers in the future and will find their ways into personality research programs.

---

[5]I did not intend my article to dismiss self-reports. I simply listed reasons that selection researchers started searching for methods other than self-reports.

REFERENCES

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin, 121*, 219–245.

Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin, 84*, 888. https://doi.org/10.1037/0033-2909.84.5.888

Allport, G. W. (1961). *Pattern and growth in personality*. New York: Holt Rinehart & Winston, Inc.

Andersen, S. M., & Chen, S. (2002). The relational self: An interpersonal social-cognitive theory. *Psychological Review, 109*, 619–645.

Aronson, E. (2010). *Not by chance alone: My life as a social psychologist*. New York: Basic Books.

Arthur, W., Jr., Woehr, D. J., & Maldegen, R. (2000). Convergent and discriminant validity of assessment center dimensions: A conceptual and empirical reexamination of the assessment center construct-related validity paradox. *Journal of Management, 26*, 813–835.

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment, 91*, 340–345.

Baard, S. K., Rench, T. A., & Kozlowski, S. W. J. (2014). Performance adaptation: A theoretical integration and review. *Journal of Management, 40*, 48–99.

Back, M. D., & Egloff, B. (2009). Yes we can! A plea for direct behavioral observation in personality research. *European Journal of Personality, 23*, 403–405.

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., Schmukle, S. C. & Egloff, B. (2010). Why are narcissists so charming at first sight? *Journal of Personality and Social Psychology, 98*, 132–145

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1–26.

Bateman, T. S., & Crant, J. M. (1993). The proactive component of organizational behavior: A measure and correlates. *Journal of Organizational Behavior, 14*, 103–118.

Baumeister, R. F. (1982). A self-presentational view of social phenomena. *Psychological Bulletin, 91*, 3–26.

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements. Whatever happened to actual behavior? *Perspectives on Psychological Science, 2,* 396–403.

Baumert, A., Halmburger, A., & Schmitt, M. (2013). Interventions against norm violations: Dispositional determinants of self-reported and real moral courage. *Personality and Social Psychology Bulletin, 39*, 1053–1068.

Beck, A. T., Freeman, A., & Davis, D. D. (2015). *Cognitive therapy of personality disorders*. New York: Guilford.

Bem, D. J. (1983). Constructing a theory of the triple typology: Some (second) thoughts on nomothetic and idiographic approaches to personality. *Journal of Personality, 51*, 566–577.

Ben-Porath, Y. S., & Tellegen, A. (1990). A place for traits in stress research. *Psychological Inquiry, 1,* 14–17.

Berry, C. M., Sackett, P. R., & Tobares, V. (2010). A meta-analysis of conditional reasoning tests of aggression. *Personnel Psychology, 63*, 361–384.

Blackie, L. E. R. & Jayawickreme, E. (2015). The example of adverse life experiences as unique situations. *European Journal of Personality, 29*, 285–386.

Blackie, L. E. R., Jayawickreme, E., Tsukayama, E., Forgeard, M. J., Roepke, A. M., & Fleeson, W. (in press). Post-traumatic growth as positive personality change: Developing a measure to assess within-person variability. *Journal of Research in Personality*.

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., Buyukcan-Tetik, A., Schwaba, T., van Scheppingen, M. A., Denissen, J. J. A., & Finkenauer, C. (2016). Stability and change in self-esteem during the transition to parenthood. *Social Psychological and Personality Science, 7,* 560–569.

Bleidorn, W., Hopwood, C. J., & Wright, A. G. (2017). *Using big data to advance personality theory.* Manuscript submitted for publication.

Blum, G., & Schmitt, M. (2017). The Nonlinear Interaction of Person and Situation (NIPS) model and its values for a psychology of situations. In *The Oxford handbook of psychological situations.* Oxford University Press.

Böckenholt, U. (2012). Modeling multiple response processes in judgment and choice. *Psychological Methods, 17,* 665–678. https://doi.org/10.1037/a0028111

Bolger, N., & Schilling, E. A. (1991). Personality and problems of everyday life: The role of neuroticism in exposure and reactivity to daily stressors. *Journal of Personality, 59,* 355–386.

Borkenau, P., Riemann, R., Angleitner, A., & Spinath, F. M. (2001). Genetic and environmental influences on observed personality: Evidence from the German observational study of adults twins. *Journal of Personality and Social Psychology, 80,* 655–668.

Borman, W. C. (1974). The rating of individuals in organizations: An alternative approach. *Organizational Behavior and Human Performance, 12,* 105–124.

Bornstein, R. F. (2009). Heisenberg, Kandinsky, and the heteromethod convergence problem: Lessons from within and beyond psychology. *Journal of Personality Assessment, 91,* 1–8.

Bouchard, T. J. (2016). Experience producing drive theory: Personality "writ large". *Personality and Individual Differences, 90,* 302–314.

Bowers, K. S. (1973). Situationism in psychology: An analysis and a critique. *Psychological Review, 80,* 307–336.

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.

Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement, 71,* 460–502. https://doi.org/10.1177/0013164410375112

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review, 62,* 193–217.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments.* Berkeley: University of California 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.

Carlson, K. D., & Ji, F. X. (2011). Citing and building on meta-analytic findings: A review and recommendations. *Organizational Research Methods, 14,* 696–717.

Carver, C. S., & Scheier, M. F. (1998). *On the self-regulation of behavior.* New York, NY: Cambridge University Press.

Caspi, A. (1998). Personality development across life course. In W. Damon (Ed.), *Handbook of child psychology* (Vol. 3, pp. 311–388). New York: Wiley.

Cesarini, D., & Visscher, P. M. (2017). Genetics and educational attainment. *NPJ Science of Learning, 2.* https://doi.org/10.1038/s41539-017-0005-6

Chabris, C. F., Lee, J. J., Cesarini, D., Benjamin, D. J., & Laibson, D. I. (2015). The fourth law of behavior genetics. *Current Directions in Psychological Science, 24,* 304–312. https://doi.org/10.1177/0963721415580430

Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology, 63,* 83–117.

Christiansen, N. D. & Robie, C. (2011). Further consideration of the usefulness of narrow trait factors. *Canadian Journal of Behavioral Sciences, 43,* 183–194.

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.* (477–497). New York (NY): Routledge.

Clark, A. (2015). *Surfing uncertainty: Prediction, action, and the embodied mind.* New York: Oxford.

Cleeton, G., & Knight, F. (1924). Validity of character judgments based on external criteria. *Journal of Applied Psychology, 8,* 215–231.

Coaster, J. A. & Christiansen, N.D. (2011). *Usefulness of O*NET work styles for determining personality-based job requirements.* Paper presented in D.L. Sandall (chair) Adequacy of O*NET Framework Work Styles at the 27th annual meeting of the Society for Industrial and Organizational Psychology, in Chicago, IL.

Coleman, J. S. & Fararo, T. J. (1992). *Rational choice theory: Advocacy and critique.* Newbury Park: Sage.

Colvin, C. R. (1993a). "Judgable" people: Personality, behavior, and competingexplanations. *Journal of Personality and Social Psychology, 64,* 861–873.

Colvin, C. R. (1993b). Childhood antecedents of young-adult judgability. *Journal of Personality, 61,* 611–635.

Connelly, B. S. (2013). Personality from the perspective of the observer: Implications for personality research and practice at work. In N. Christiansen & R. Tett (Eds). *Handbook of personality at work.* New York, Taylor-Francis/Routledge Press.

Connelly, B. S., & Ones, D. S. (2010). Another perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin, 136,* 1092–1122.

Cooper, W. H., & Withey, M. J. (2009). The strong situation hypothesis. *Personality and Social Psychology Review, 13,* 62–72.

Corr, P. J. (2009). Behaviour functions in personality psychology. *European Journal of Personality, 23,* 403–435.

Costa, P.T., Jr., & McCrae, R. R. (1980). Still stable after all these years: Personality as a key to some issues in adulthood and old age. In P. B. Baltes & O. G. Brim, Jr. (Eds.), *Lifespandevelopment and behavior* (Vol. 3, pp. 65–102). New York: Academic Press.

Creswell, K. G., Bachrach, R. L., Wright, A. G., Pinto, A., & Ansell, E. (2016). Predicting problematic alcohol use with the DSM-5 alternative model of personality pathology. *Personality disorders, 7,* 103–111.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12,* 671–684.

Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1963). Theory of generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology, 16,* 137–163.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles.* New York: Wiley.

Culbertson, S. S., Weyhrauch, W. S., & Huffcutt, A. I. (2017). A tale of two formats: Direct comparison of matching situational and behavior description interview questions. *Human Resource Management Review, 27,* 167–177. https://doi.org/10.1016/j.hrmr.2016.09.009

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 Fruyt, F., & Salgado, J. F. (2003). Editorial: Personality and IWO applications: Introducing personality at work. *European Journal of Personality, 17,* S1-S3. https://doi.org/10.1002/per.480

De Fruyt, F., & Van Leeuwen, K. (2014). Advancements in the field of personality development. *Journal of Adolescence, 37,* 763–769. https://doi.org/10.1016/j.adolescence.2014.04.009

De Kock, F., Lievens, F., & Born, M. Ph. (2015). An in-depth look at dispositional reasoning and interviewer accuracy. *Human Performance, 28,* 199–221.

De Kock, F., Lievens, F., & Born, M. Ph. (2017). A closer look at the measurement of dispositional reasoning: Dimensionality and invariance across assessor groups. *International Journal of Selection and Assessment*, 25, 240–252.

DeYoung, C. G. (2015). Cybernetic Big Five theory. *Journal of Research in Personality*, 56, 33–58.

Dunlop, W. L. (2015). Contextualized personality, beyond traits. *European Journal of Personality, 29,* 310–325. https://doi.org/10.1002/per.1995

Engelhard, I. M., Lommen, M. J., & Sijbrandij, M. (2015). Changing for better or worse? Posttraumatic growth reported by soldiers deployed to Iraq. *Clinical Psychological Science, 3,* 789–796.

Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology, 37,* 1097. https://doi.org/10.1037/0022-3514.37.7.1097

Eysenck, H. J. (1997). Personality and experimental psychology: The unification of psychology and the possibility of a paradigm. *Journal of Personality and Social Psychology, 73,* 1224–1237.

Few, L. R., Miller, J. D., Rothbaum, A. O., Meller, S., Maples, J., Terry, D. P., … Mackillop, J. (2013). Examination of the Section Iii Dsm-5 diagnostic system for personality disorders in an outpatient clinical sample. *Journal of abnormal psychology*, 122, 1057–1069.

Fishbein, M., & Ajzen, I. (1974). Attitudes toward objects as predictors of single and multiple behavioral criteria. *Psychological Review, 81,* 59–74.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley Pub.

Fiske, S. T., & Macrae, C. N. (Eds.). (2012). *The Sage handbook of social cognition*. Thousand Oaks, CA: Sage Publications.

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. (2007). Situation-based contingencies underlying trait-content manifestation in behavior. *Journal of Personality, 75,* 825–862.

Fleeson, W. (2014). Four ways of (not) being real and whether they are essential for post-traumatic growth. *European Journal of Psychology, 28*, 336–337.

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.

Fournier, M. A., Moskowitz, D. S., & Zuroff, D. C. (2008). Integrating dispositions, signatures, and the interpersonal domain. *Journal of Personality and Social Psychology*, 94, 531–545.

Fournier, M. A., Moskowitz, D. S., & Zuroff, D. C. (2009). The interpersonal signature. *Journal of Research in Personality, 43*, 155–162.

Frazier, P., Coyne, J. & Tennen, H. (2014). Post-traumatic growth: A call for less, but better, research. *European Journal of Psychology, 28*, 337–338.

Frederiksen, N. (1972). Toward a taxonomy of situations. *American Psychologist, 27,* 114–123.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11, 127–138.

Funder, D. C. (1991). Global traits: A neo-Allportian approach to personality. *Psychological Science, 2,* 31–39.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102, 652–670.

Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception*. San Diego: Academic Press.

Funder, D. C. (2006). Towards a resolution of the personality triad: Persons, situations, and behaviors. *Journal of Research in Personality, 40,* 21–34.

Funder, D. C. (2016). Taking situations seriously: The situation construal model and the riverside situational Q-sort. *Current Directions in Psychological Science, 25,* 203–208.

Funder, D. C., & Colvin, C. R. (1991). Explorations in behavioral consistency: Properties of persons, situations, and behaviors. *Journal of Personality and Social Psychology, 60,* 773–794. https://doi.org/10.1037/0022-3514.60.5.773

Furr, R. M. (2009). The study of behaviour in personality psychology: Meaning, importance and measurement. *European Journal of Personality*, 23, 437–453.

Geukes, K., Nestler, S., Hutteman, R., Küfner, A. C. P., & Back, M. D. (2017). 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*, 69, 124–138.

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.

Grove, W. M., & Tellegen, A. (1991). Problems in the classification of personality disorders. *Journal of Personality Disorders, 5,* 31–41.

Halberstadt, A. G. (1983). Family expressiveness styles and nonverbal communicationskills. *Journal of Nonverbal Behavior*, 8, 14–26.

Handler, C. (2013). Foreword. In M. S. Fetzer & K.A. Tuzinski (Eds.), *Simulations for personnel selection* (pp. v–ix). New York: Springer.

Herde, C. N., & Lievens, F. (2016). *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.

Highhouse, S., & Harris, M. M. (1993). The measurement of assessment center situations: Bem's template matching technique for examining exercise similarity. *Journal of Applied Social Psychology, 23,* 140–155.

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., & Woehr, D. J. (2009). Disentangling the meaning of multisource performance rating source and dimension factors. *Personnel Psychology, 62,* 735–765.

Hofmann, W., Gschwendner, T., Nosek, B. A., & Schmitt, M. (2005). What moderates implicit—Explicit consistency? *European Review of Social Psychology, 16,* 335–390.

Hogan, R. (1996). A socioanalytic interpretation of the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality* (pp. 163–179). New York: Guilford.

Hohwy, J. (2013). *The predictive mind*. New York: Oxford University Press.

Hopwood, C. J. & Bornstein, R. F. (2014). *Multimethod clinical assessment*. New York, NY: Guilford.

Hopwood, C. J., Malone, J. C., Ansell, E. B., Sanislow, C. A., Grilo, C. M., McGlashan, T. H., … & Gunderson, J. G. (2011). Personality assessment in DSM-5: Empirical support for rating severity, style, and traits. *Journal of Personality Disorders*, 25, 305–320.

Horstmann, K. T., Rauthmann, J. F., & Sherman, R. A. (2017). The measurement of situational influences. In V. Zeigler-Hill and T. K. Shackelford (Eds.), *The SAGE handbook of personality and individual differences*.

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75,* 581–595.

Human, L. J., & Biesanz, J. C. (2011). Target adjustment and self-other agreement: Utilizing trait observability to disentangle judgeability and self-knowledge. *Journal of Personality and Social Psychology*, 101, 202–216.

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.

Human, L. J., Biesanz, J. C., Parisotto, K. L., & Dunn, E. W. (2012). Your best self helps reveal your true self: Positive self-presentation leads to more accurate personalityimpressions. *Social Psychological and Personality Science, 3*, 23–30.

Human, L. J., Sandstrom, G. M., Biesanz, J. C., & Dunn, E. W. (2013). Accurate first impressions leave a lasting impression: The long-term effects of distinctive self-otheragreement on relationship development. *Social Psychological and Personality Science, 4*, 395–402.

Human, L. J., Biesanz, J. C., Finseth, S. M., Pierce, B., & Le, M. (2014). To thine own self rue: Psychological adjustment promotes judgeability via personality–behavior congruence. *Journal of Personality and Social Psychology, 106*, 286–303.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting for error and bias in research findings*. Thousand Oaks, CA: Sage.

Ickes, W., Snyder, M., & Garcia, S. (1997). Personality influences on the choice of situations. In R. Hogan, J. Johnson, & S. Briggs (Eds.) *Handbook of personality psychology* (pp. 165–195). San Diego, CA: Academic Press.

Infurna, F. J., & Luthar, S. S. (2016). Resilience to major life stressors is not as common as thought. *Perspectives on Psychological Science, 11*, 175–194.

International Taskforce on Assessment Center Guidelines (2015). Guidelines and ethical considerations for assessment center operations. *Journal of Management, 41,* 1244–1273.

James, L. R. (1998). Measurement of personality via conditional reasoning. *Organizational Research Methods, 2*, 131–163.

James, L. R., & McIntyre, M. D. (2000). *Conditional reasoning test of aggression test manual*. Knoxville, TN: Innovative Assessment Technology.

Janz, T. (1982). Initial comparisons of patterned behavior description interviews versus unstructured interviews. *Journal of Applied Psychology, 67*, 577–580. https://doi.org/10.1037/0021-9010.67.5.577

Jayawickreme, E. & Blackie, L. E. R. (2014). Post-traumatic growth as positive personality change: Evidence, controversies and future directions. *European Journal of Personality, 28,*312–31

Johnson, W. (2010). Extending and testing Tom Bouchard's experience producing drive theory. *Personality and Individual Differences, 49*, 296–301.

Jones, A. B., Brown, N. A., Serfass, D. G., & Sherman, R. A. (2014). Traumatic experiences yield changes in the situation experience. *European Journal of Personality, 28*, 339–340.

Kammrath, L. K., Mendoza-Denton, R., & Mischel, W. (2005). Incorporating if…then…personality signatures in person perception: Beyond the person-situation dichotomy. *Journal of Personality and Social Psychology, 88*, 605–618.

Kelley, H. H., Holmes, J. G., Kerr, N. L., Reis, H. T., Rusbult, C. E., & Van Lange, P. A. M. (2003). *An atlas of interpersonal situations*. Cambridge: Cambridge University Press.

Kenny, D. A., & West, T. V. (2008). Self-perception as interpersonal perception. In J. V. Wood, A. Tesser, & J. G. Holmes (Eds.), *The self and social relationships*. (pp. 119–138). New York: Psychology Press.

Klehe, U.-C., & Latham, G. (2006). What would you do–really or ideally? Constructs underlying the behavior description interview and the situational interview in predicting typical versus maximum performance. *Human Performance, 19*, 357–382. https://doi.org/10.1207/s15327043hup1904_3

Kleinmann, M. (1997). Transparenz der Anforderungsdimensionen: Ein moderator der Konstrukt und Kriteriumsvalidität [transparency of the required dimensions: A moderator of assessment centers' construct and criterion validity]. *Zeitschrift für Arbeits- und Organisationspsychologie, 41,* 171–181.

Klirs, E. G., & Revelle, W. (1986). Predicting variability from perceived situational similarity. *Journal of Research in Personality, 20*, 34–50.

Kluemper, D., McLarty, B., & Bing, M. (2015). Acquaintance ratings of conscientiousness, agreeableness, and emotional stability: Incremental validity beyond and interactive effects with self-reports in the prediction of workplace deviance. *Journal of Applied Psychology, 100*, 237–248.

Kluemper, D. H., McLarty, B. D., & Bing, M. N. (2017). *The validity of friend, family, and coworker ratings of personality*. Paper presented at the 32nd annual meeting of the Society for Industrial and Organizational Psychology Conference held in Orlando, FL.

Knafo, A., & Jaffee, S. R. (2013). Gene–environment correlation in developmental psychopathology. *Development and Psychopathology, 25*, 1–6.

Kotov, R., Gamez, W., Schidt, F., & Watson, D. (2010). Linking big personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological Bulletin, 136*, 768–821.

Krumm, S., Lievens, F., Hüffmeier, J., Lipnevich, A. A., Bendels, H., & Hertel, G. (2015). How "situational" is judgment in situational judgment tests? *Journal of Applied Psychology, 100*, 399–416. https://doi.org/10.1037/a0037674

Krumm, S., Hüffmeier, J., & Lievens, F. (2017). Experimental test validation. *European Journal of Psychological Assessment.* https://doi.org/10.1027/1015-5759/a000393

Kuncel, N. R., & Sackett, P. R. (2014). Resolving the assessment center construct validity problem (as we know it). *Journal of Applied Psychology, 99*, 38–47. https://doi.org/10.1037/a0034147

Lance, C. E., Noble, C. L., & Scullen, S. E. (2002). A critique of the correlated trait-correlated method and correlated uniqueness models for multitrait-multimethod data. *Psychological Methods, 7*, 228–244. https://doi.org/10.1037//1082-989X.7.2.228

Lance, C. E., Dawson, B., Birkelbach, D., & Hoffman, B. J. (2010). Method effects, measurement error, and substantive conclusions. *Organizational Research Methods, 13*, 435–455.

Lang, J. W. B. (2014). A dynamic Thurstonian item response theory of motive expression in the picture story exercise: Solving the internal consistency paradox of the PSE. *Psychological Review, 121*, 481–500. https://doi.org/10.1037/a0037011

LaPiere, R. T. (1934). Attitudes versus actions. *Social Forces, 13*, 230–237.

Lawson, C., & MacLeod, C. (1999). Depression and the interpretation of ambiguity. *Behaviour Research and Therapy, 37*, 463–474.

Leary, M. R. (1995). *Self-presentation: Impression management and interpersonal behavior*. Brown & Benchmark Publishers.

Leary, T., & Coffey, H. S. (1955). Interpersonal diagnosis: Some problems of methodology and validation. *Journal of Abnormal and Social Psychology, 50*, 110–124.

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.

Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology, 67*, 241–293. https://doi.org/10.1111/peps.12052

Lewin, K. (1943). Defining the "field at a given time". *Psychological Review, 50*, 292–310.

Lievens, F. (2001). Assessors and use of assessment center dimensions: A fresh look at a troubling issue. *Journal of Organizational Behavior, 65*, 1–19.

Lievens, F. (2017). Assessing personality-situation interplay in personnel selection: Towards more integration into personality research. *European Journal of Personality,*

Lievens, F. (in press). Construct-driven SJTs: Towards an agenda for future research. *International Journal of Testing*.

Lievens, F., & Conway, J. M. (2001). Dimension and exercise variance in assessment center scores: A large-scale evaluation of multitrait-multimethod studies. *Journal of Applied Psychology, 86*, 1202–1222.

Lievens, F., & Motowidlo, S. J. (2016). Situational judgment tests: From measures of situational judgment to measures of general domain knowledge. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 9*, 3–22. https://doi.org/10.1017/iop.2015.71

Lievens, F. & Peeters, H. (2008). Impact of elaboration on responding to situational judgment test items. *International Journal of Selection and Assessment, 16*, 345–355.

Lievens, F., & Sackett, P. R. (2012). The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *Journal of Applied Psychology, 97*, 460–468.

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. https://doi.org/10.1037/apl0000160

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., Schollaert, E., & Keen, G. (2015). The interplay of elicitation and evaluation of trait-expressive behavior: Evidence in assessment center exercises. *Journal of Applied Psychology, 100*, 1169–1188.

Lilienfeld, S. O. (2013). Is psychopathy a syndrome? Commentary on Marcus, Fulton, and Edens. *Personality Disorders: Theory, Research, and Treatment, 4*, 85–86.

Lönnqvist, J. E., Paunonen, S., Verkasalo, M., Leikas, S., Tuulio-Henriksson, A., & Lönnqvist, J. (2007). Personality characteristics of research volunteers. *European Journal of Personality, 21*, 1017–1030.

Lord, R. G. (1985). Accuracy in behavioral measurement: An alternative definition based on raters' cognitive schema and signal detection theory. *Journal of Applied Psychology, 70*, 66–71.

Lowman, G., Wood, D., & Harms, P. D. (2017). *Assertiveness and power in the workplace: A functional field approach*. University of Alabama, Tuscaloosa, AL.

Lucas, R. E., Clark, A. E., Georgellis, Y., & Diener, E. (2003). Reexamining adaptation and the set point model of happiness: Reactions to changes in marital status. *Journal of Personality and Social Psychology, 84*, 527–539.

MacDonald, K. (1995). Evolution, the five-factor model, and levels of personality. *Journal of Personality, 63*, 525–567.

MacLane, C. N. & Cucina, J. M. (2015). Generalization of cognitive and noncognitive validities across personality-based job families. *International Journal of Selection and Assessment 23*, 316–331

Magnusson, D. (1981). *Toward a psychology of situations: An interactional perspective*. Hillsdale, NJ: Erlbaum.

Mancini, A. D., Littleton, H. L., & Grills, A. E. (2016). Can people benefit from acute stress? Social support, psychological improvement, and resilience after the Virginia Tech campus shootings. *Clinical Psychological Science, 4*, 401–417.

Markon, K. E., Krueger, R. F., & Watson, D. (2005). Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. *Journal of Personality and Social Psychology, 88*, 139–157.

McAbee, S. T., & Connelly, B. S. (2016). A multi-rater framework for studying personality: The trait-reputation-identity model. *Psychological Review, 123*, 569–591. https://doi.org/10.1037/rev0000035

McAdams, D. P. (1995). What do we know when we know a person? *Journal of Personality, 63*, 365–396. https://doi.org/10.1111/j.1467-6494.1995.tb00500.x

McCrae, R. R. (2015). A more nuanced view of reliability: Specificity in the trait hierarchy. *Personality and Social Psychology Review, 19*, 97–112. https://doi.org/10.1177/1088868314541857

McCrae, R. R., & Weiss, A. (2007). Observer ratings of personality. In R. W. Robins, R. C. Fraley, & R. Krueger (Eds.), *Handbook of research methods in personality psychology*. (pp. 259–272). New York, NY: Guilford Press.

McCrae, R. R., Costa, P. T., Jr., & Martin, T. A. (2005). The NEO-PI-e: A more readable revised NEO personality inventory. *Journal of Personality Assessment, 84*, 261–270.

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). *Use of situational judgment tests to predict job performance: A clarification of the literature*. American Psychological Association.

McDaniel, M. A., List, S. K., & Kepes, S. (2016). The "hot mess" of situational judgment test construct validity and other issues. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 9*, 47–51.

McFarland, L. Y. N. N., Yun, G. J., Harold, C. M., Viera, L., & Moore, L. G. (2005). An examination of impression management use and effectiveness across assessment center exercises: The role of competency demands. *Personnel Psychology, 58*, 949–980.

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.

Meriac, J. P., Hoffman, B. J., & Woehr, D. J. (2014). A conceptual and empirical review of the structure of assessment center dimensions. *Journal of Management, 40*, 1269–1296. https://doi.org/10.1177/0149206314522299

Meyer, R. D., & Dalal, R. S. (2009). Situational strength as a means of conceptualizing context. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 2*, 99–102.

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.

Miller, K. E., & Rasmussen, A. (2014). War experiences, daily stressors and mental health five years on: Elaborations and future directions. *Intervention, 12*, 33–42.

Mischel, W. (1968). *Personality and assessment*. New York: John Wiley & Sons.

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.

Miskewicz, K., Fleeson, W., Mayfield Arnold, E., Law, M.K., Mneimne, M., & Furr, R.M. (2015). A contingency-oriented approach to understanding borderline personality disorder: Situational triggers and symptoms. *Journal of Personality Disorders, 29*, 486–502. https://doi.org/10.1521/pedi.2015.29.4.486

Moore, C., Lee, S. Y., Kim, K., & Cable, D. M. (2017). The advantage of being oneself: The role of applicant self-verification in organizational hiring decisions. *Journal of Applied Psychology*. In Press.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology, 60*, 683–729.

Morse, P. J., Sauerberger, K. S., Todd, E., & Funder, D. (2015). Relationships among personality, situational construal and social outcomes. *European Journal of Personality, 29*, 97–106. https://doi.org/10.1002/per.1987

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

Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006). Implicit policies about relations between personality traits and behavioral effectiveness in situational judgment items. *Journal of Applied Psychology*, *91*, 749–761.

Mõttus, R. (2016). Towards more rigorous personality trait–outcome research. *European Journal of Personality, 30*, 292–303. https://doi.org/10.1002/per.2041

Mõttus, R., Kandler, C., Bleidorn, W., Riemann, R., & McCrae, R. R. (2017a). Personality traits below facets: The consensual validity, longitudinal stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology, 112*, 474–490. https://doi.org/10.1037/pspp0000100

Mõttus, R., Bates, T. C., Condon, D. M., Mroczek, D. K., & Revelle, W. (2017b). *Your personality data can do more: Items provide leverage for explaining the variance and covariance of life outcomes.* Retrieved from osf.io/preprints/psyarxiv/4q9gv

Murphy, N. A. (2007). Appearing smart: The impression management of intelligence, person perception accuracy, and behavior in social interaction. *Personality and Social Psychology Bulletin*, *33*, 325–339.

Murray, H. A. (1938). *Explorations in personality: A clinical and experimental study of fifty men of college age*. Oxford: Oxford University Press.

Mussel, P. (2007). *Die Konstruktvalidität des Multimodalen interviews [construct validity of the multimodal employment interview]*. Berlin: dissertation.de.

Mussel, P., Gatzka, T., & Hewig, J. (2016). Situational judgment tests as an alternative measure for personality assessment. *European Journal of Psychological Assessment*, https://doi.org/10.1027/1015-5759/a000346

Nesselroade, J. R., & Boker, S. M. (1994). Assessing constancy and change. In T. Heatherton & J. Weinberger (Eds.), *Can personality change?* (pp. 121–147). Washington, DC: American Psychological Association.

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.

Ones, D. S., & Viswesvaran, C. (1996). Bandwidth-fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior, 17*, 609–626.

Parrigon, S., Woo, S. E., Tay, L., & Wang, T. W. (2016). CAPTION-ing the situation: A lexically-derived taxonomy of psychological situation characteristics. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspp0000111

Patrick, C. J. (in press). Psychopathy as masked pathology. In C.J. Patrick (Ed.), *Handbook of psychopathy*. New York: Guilford.

Pincus, A. L. (2011). Some comments on nomology, diagnostic process, and narcissistic personality disorder in the DSM-5 proposal for personality and personality disorders. *Personality Disorders*, *2*, 41–53.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879–903.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*, 539–569.

Popple, A., & Levi, D. M. (2000). Wundt versus Galton – two approaches to gathering psychophysical measurements. *Perception*, *29*, 379–381.

Poropat, A. & Corr, P. J (2015). Thinking bigger: The Cronbachian paradigm and personality theory integration. *Journal of Research in Personality*, *56*, 59–69.

Rauthmann, J. F. (2015). Structuring situational information: A road map of the many pathways to different situational taxonomies. *European Psychologist, 20,* 176–189.

Rauthmann, J. F. (2017). *Situation management strategies dendrogram (Figure)*. https://doi.org/10.17605/OSF.IO/D3562

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., Jones, A. B., & Sherman, R. A. (2016). Directionality of person-situation transactions: Are there spill-overs among and between situation experiences and personality states? *Personality and Social Psychology Bulletin, 42*, 893–909.

Rauthmann, J. F., & Sherman, R. A. (2016a). Measuring the situational eight DIAMONDS characteristics of situations: An optimization of the RSQ-8 to the S8*. *European Journal of Psychological Assessment, 32,* 155–164

Rauthmann, J. F., & Sherman, R. A. (2016b). Ultra-brief measures for the situational eight DIAMONDS domains. *European Journal of Psychological Assessment*, *32*, 165–174.

Rauthmann, J. F., & Sherman, R. A. (2016c). Situation change: Stability and change of situation variables between and within persons. *Frontiers in Psychology*, *6*, 1938. https://doi.org/10.3389/fpsyg.2015.01938

Rauthmann, J. F., & Sherman, R. A. (2017). Normative and distinctive accuracy in situation perceptions: Magnitude and personality correlates. *Social Psychological and Personality Science*, 1–12.

Rauthmann, J. F., Sherman, R. A., & Funder, D. C. (2015). Principles of situation research: Towards a better understanding of psychological situations. *European Journal of Personality*, *29*, 363–381.

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.

Raymark, P. H., & van Iddekinge, C. H. (2013). Assessing personality in selection interviews. In N. D. Christiansen & R. P. Tett (Eds.), *Handbook of personality at work*. New York, NY: Routledge.

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience?. *Personality and Social Psychology Review*, *8*, 364–382.

Reis, H. T. (2008). Reinvigorating the concept of situation in social psychology. *Personality and Social Psychology Review*, *12*, 311–329.

Revelle, W., & Condon, D. M. (2015). A model for personality at three levels. *Journal of Research in Personality*, *56*, 70–81. https://doi.org/10.1016/j.jrp.2014.12.006

Reynolds, K. J., Turner, J.C., Branscombe, N. R., Mavor, K.I., Bizumic, B., & Subasic, E. (2010). Interactionism in personality and social psychology: An integrated approach to understanding the mind and behaviour. *European Journal of Personality, 24*, 458–482.

Robinson, M. D., & Clore, G. L. (2002). Episodic and semantic knowledge in emotional self-report: Evidence for two judgment processes. *Journal of Personality and Social Psychology*, *83*, 198–215.

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.

Rogers, K. H. (2015). *The "good dyad": Examining the impact of personality and behavior on dyadic accuracy in first impressions*. (unpublished doctoral dissertation). University of British Columbia, Vancouver, Canada.

Rogers, K. H., & Biesanz, J. C. (January, 2016). *The good judge revisited: Examining the role of the good target. Poster presented at Society for Personality and Social Psychology*, San Diego, CA.

Rozin, P. (2009). What kind of empirical research should we publish, fund, and reward?: A different perspective. *Perspectives on Psychological Science*, *4*, 435–439.

Sackett, P. R., & Dreher, G. F. (1982). Constructs and assessment center dimensions: Some troubling empirical findings. *Journal of Applied Psychology, 67,* 401–410.

Sackett, P. R., Zedeck, S., & Fogli, L. (1988). Relations between measures of typical and maximum job performance. *Journal of Applied Psychology, 73,* 482–486.

Salgado, J. F. (1997). The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology, 82,* 30–43.

Sauerberger, K. S., & Funder, D. C. (2016). Behavioral change and consistency across contexts. *Journal of Research in Personality.* https://doi.org/10.1016/j.jrp.2016.04.007

Scarr, S., & McCartney, K. (1983). How people make their own environments: A theory of genotype→environment effects. *Child Development, 54,* 424–435.

Schlenker, B. R., & Pontari, B. A. (2000). The strategic control of information: Impression management and self-presentation in daily life. In A. Tesser, R. B. Felson, & J. M. Suls (Eds.), *Psychological perspectives on self and identity* (pp. 199–232). Washington, DC: American Psychological Association.

Schmid Mast, M., Gatica-Perez, D., Frauendorfer, D., Nguyen, L., & Choudhury, T. (2015). Social sensing for psychology: Automated interpersonal behavior assessment. *Current Directions in Psychological Science, 24,* 87–92.

Schmidt, F. L., Oh, I. S., & Shaffer, J. A. (2016). *The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 100 years of research findings.* Retrieved from Website. https://www.researchgate.net/publication/309203898

Schneider, J. R., & Schmitt, N. (1992). An exercise design approach to understanding assessment center dimension and exercise constructs. *Journal of Applied Psychology, 77,* 32–41.

Schollaert, E., & Lievens, F. (2011). The use of role-player prompts in assessment center exercises. *International Journal of Selection and Assessment, 19,* 190–197.

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.

Sharp, C., Wright, A. G. C., Fowler, J. C., Freuh, C., Allen, J. G., Oldham, J., & Clark, L. A. (2015). The structure of personality pathology: Both general ('g') and specific ('s') factors? *Journal of Abnormal Psychology, 124,* 387–398.

Sheldon, K. M., & Kasser, T. (1995). Coherence and congruence: Two aspects of personality integration. *Journal of Personality and Social Psychology, 68,* 531–543.

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. G., & 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.

Spearman, C. (1961). C. Spearman. In C. Murchison (Ed.), *A history of psychology in autobiography* (Vol. 1, pp. 299–333). New York: Russell and Russell (Original work published 1930).

Specht, J., Bleidorn, W., Denissen, J. J. A., Hennecke, M., Hutteman, R., Kandler, C., … Zimmermann, J. (2014). What drives adult personality development? A comparison of theoretical perspectives and empirical evidence. *European Journal of Personality, 28,* 216–230. https://doi.org/10.1002/per.1966

Speer, A., Christiansen, N. D., Goffin, R. D., & Goff, M. (2014). Situational bandwidth and the criterion-related validity of assessment center ratings: Is cross-exercise convergence always desirable? *Journal of Applied Psychology, 100,* 1–14.

Speer, A., Christiansen, N. D., & Honts, C. (2015). Assessment of personality through behavioral observations in work simulations. *Personnel Assessment and Decisions, 1,* 43–56.

Suzuki, T., Samuel, D. B., Pahlen, S., & Krueger, R. F. (2015). Dsm-5 alternative personality disorder model traits as

maladaptive extreme variants of the five-factor model: An item-response theory analysis. *Journal of abnormal psychology, 124,* 343–354.

Swann, W. B., Johnson, R. E., & Bosson, J. K. (2009). Identity negotiation at work. *Research in Organizational Behavior, 29,* 81–109.

Tackett, J. L., Herzhoff, K., Kushner, S. C., & Rule, N. (2016). Thin slices of child personality: Perceptual, situational, and behavioral contributions. *Journal of Personality and Social Psychology, 110,* 150–166. https://doi.org/10.1037/pspp0000044

Tackett, J. L., Lang, J. W. B., Markon, K. E., & Herzhoff, K. (under review). *A correlated traits correlated methods model for thin-slice child personality assessment. Submitted for Publication.*

Taylor, S. E. (2006). Tend and befriend: Biobehavioral bases of affiliation under stress. *Current Directions in Psychological Science, 15,* 273–277. https://doi.org/10.1111/j.1467-8721.2006.00451.x

Tedeschi, R. G., & Calhoun, L. G. (2004) Posttraumatic growth: Conceptual foundations and empirical evidence. *Psychological Inquiry 15,* 1–18. https://doi.org/10.1207/s15327965pli1501_01.

Tennen, H., & Affleck, G. (1998). Personality and transformation in the face of adversity. In Tedeschi, R. G., Park, C. & Calhoun, L. G. (Eds.), *Posttraumatic growth: Positive changes in the aftermath of crisis* (pp. 65–98). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Tellegen, A. (1981). Practicing the two disciplines for relaxation and enlightenment: Comment on "Role of the feedback signal in electromyograph biofeedback: The relevance of attention" by Qualls and Sheehan. *Journal of Experimental Psychology: General, 110,* 217–226.

Tellegen, A. (1991). Personality traits: Issues of definition, evidence, and assessment. In Cicchetti, D. & Grove, W. M. (Eds.), *Thinking clearly about psychology: Essays in honor of Paul E. Meehl* (pp. 10–35). Minneapolis, MN: University of Minnesota Press.

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., & Christiansen, N. D. (2007). Personality tests at the crossroads: A response to Morgeson, Campion, Dipboye Hollenbeck, Murphy, and Schmitt (2007). *Personnel Psychology, 60,* 967–993.

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.

Tett, R. P., Jackson, D. N., Rothstein, M., & Reddon, J. R. (1999). Meta-analysis of bi-directional relations in personality-job performance research, *Human Performance, 12,* 1–29.

Tett, R. P., Simonet, D. V., Walser, B., & Brown, C. (2013). Trait activation theory: Applications, developments, and implications for person-workplace fit. In N. D. Christiansen & R. P. Tett (eds.) *Handbook of personality at work* (pp. 71–100). New York: Routledge.

Tett, R. P., Hundley N. A., & Christiansen, N. D. (2017). Meta-analysis and the myth of generalizability. *Industrial and Organizational Psychology: Perspectives on Science and Practice,* in press.

Thornton, G. C., Rupp, D. E., & Hoffman, B. (2015). *Assessment center perspectives for talent management strategies.* New York: Routlege.

Thornton, G. C., Hanson, R. M., & Rupp, D. E. (2017). *Developing organizational simulations: A guide for practitioners, students, and researchers.* New York, NY: Routledge.

Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review, 93,* 239–257.

Tupes, E. C. (1957). *Relationships between behavior trait ratings by peers and later office performance of USAF officer candidate school graduates.* Lackland Air Force Base, Texas: Air Force Personnel and Training Research Center.

Tupes, E. C., & Christal, R. C. (1958). *Stability of personality trait rating factors obtained under diverse conditions USAF Wright*

Air Development Center Technical Note. Lackland Air Force Base: United States Air Force.

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. https://doi.org/10.1037/0021-9010.90.3.536

van Scheppingen, M. A., Jackson, J. J., Specht, J., Hutteman, R., Denissen, J. J. A., & Bleidorn, W. (2016). Personality development during the transition to parenthood: A test of social investment theory. *Social Psychological and Personality Science, 7,* 452–462.

van Scheppingen, M. A., Denissen, J. J. A., Chung, J. M., Tambs, K., & Bleidorn, W. (2017). Self Esteem and Relationship Satisfaction During the Transition to Motherhood. *Journal of Personality and Social Psychology*. Preprint available at https://osf.io/xv4np/

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.

Verbeke, L., & De Clercq, B. (2017). *The development of a situational judgement tool (SJT) for personality disorders*. Contribution to symposium entitled: "'New insights and trends in the assessment of personality pathology" (Convener: Verbeke, L.) at the XVth Conference of the International Society for the Study of Personality Disorders (ISSPD), 2017, September 25-28, Heidelberg, Germany.

Wachtel, P. (1973). Psychodynamics, behavior therapy, and the implacable experimenter: An inquiry into the consistency of personality. *Journal of Abnormal Psychology, 82*, 324–334.

Wagerman, S. A., & Funder, D. C. (2009). Situations. In P. J. Corr & G. Matthews (Eds.), *Cambridge handbook of personality psychology* (pp. 27–42). Cambridge, England: Cambridge University Press.

Walton, K. E., Roberts, B. W., Krueger, R. F., Blonigen, D. M., & Hicks, B. M. (2008). Capturing abnormal personality with normal personality inventories: An item response theory approach. *Journal of personality*, *76*, 1623–1648.

Whitener, E. M. (1990). Confusion of confidence intervals and credibility intervals in meta-analysis. *Journal of Applied Psychology, 75*, 315–321.

Wicker, A. W. (1969). Attitudes versus actions: The relationship of verbal and overt behavioural responses to attitude objects. *Journal of Social Issues, 25*, 41–78.

Wilson, S., Stroud, C. B., & Durbin, C. E. (2017). Interpersonal dysfunction in personality disorders: A meta-analytic review. *Psychological Bulletin, 143*, 677–734.

Wood, D., Gardner, M. H., & Harms, P. D. (2015). How functionalist and process approaches to behavior can explain trait covariation. *Psychological Review*, *122*, 84–111.

Wood, D., Lowman, G., Harms, P. D., & Spain, S. M. (in press). *Using functional fields to formally represent the meaning and logic of actions: A worked example using 'dark' actions*.

Wood, D., Spain, S. M., & Harms, P. D. (in press). Functional approaches to representing the interplay of situations, traits, and behavior. In D. C. Funder, R. A. Sherman, & J. F. Rauthmann, *Handbook of psychological situations*. New York: Oxford.

Wright, S. (1934). The method of path coefficients. *The Annals of Mathematical Statistics*, *5*, 161–215.

Wright, A. G. C. (2011). Quantitative and qualitative distinctions in personality disorder. *Journal of Personality Assessment, 93*, 370–379.

Wright, A. G. C., & Hopwood, C. J. (2016). Advancing the assessment of dynamic psychological processes. *Assessment*, *23*, 399–403.

Wright, A. G. C., & Simms, L. J. (2014). On the structure of personality disorder traits: Conjoint analyses of the CAT-PD, PID-5, and NEO-PI-3 trait models. *Personality Disorders: Theory, Research, and Treatment, 5*, 43–54.

Wright, A. G. C., & Simms, L. J. (2016). Stability and fluctuation of personality disorder features in daily life. *Journal of Abnormal Psychology. 125*, 641–656.

Wright, A. G. C. & Zimmermann, J. (2015). At the nexus of science and practice: Answering basic clinical questions in personality disorder assessment and diagnosis with quantitative modeling techniques. In S. Huprich (Ed.), *Personality disorders: Toward theoretical and empirical integration in diagnosis and assessment* (pp. 109–144). Washington, DC: American Psychological Association.

Wright, A. G. C., Thomas, K. M., Hopwood, C. J., Markon, K. E., Pincus, A. L. & Krueger, R. F. (2012). The hierarchical structure of DSM-5 pathological personality traits. *Journal of Abnormal Psychology, 121*, 951–957.

Wright, A. G. C., Stepp, S. D., Scott, L. N., Hallquist, M. N., Beeney, J. E., Lazarus, S., & Pilkonis, P. A. (in press). The effect of pathological narcissism on interpersonal and affective processes in social interactions. *Journal of Abnormal Psychology*.

Yarkoni, T., & Westfall, J. (in press). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science, in press*.

Zaki, J., Bolger, N., & Ochsner, K. (2008). It takes two: The interpersonal nature of empathic accuracy. *Psychological Science*, *19*, 399–404.

Zettler, I., Lang, J. W. B., Hülsheger, U. R., & Hilbig, B. E. (2016). Dissociating indifferent, directional, and extreme responding in personality data: Applying the three-process model to self- and observer reports. *Journal of Personality*, *84*, 461–472. https://doi.org/10.1111/jopy.12172

Ziegler, M., & Brunner, M. (2016). Test standards and psychometric modeling. In A. A. Lipnevich, F. Preckel, & R. Roberts (Eds.), *Psychosocial skills and school systems in the 21st century* (pp. 29–55). Göttingen: Springer.

Ziegler, M., & Horstmann, K. (2015). Discovering the second side of the coin: Integrating situational perception into psychological assessment. *European Journal of Personality Assessment, 31*, 69–74.

Ziegler, M., MacCann, C., & Roberts, R. D. (2011). *New perspectives on faking in personality assessments*. New York: Oxford Press.

Zimmerman, R. D., del Triana, M. C., & Barrick, M. R. (2010). Predictive criterion-related validity of observer ratings of personality and job-related competencies using multiple raters and multiple performance criteria. *Human Performance*, *23*, 361–378.