

This preprint version may not exactly replicate the final version published in *Journal of Applied Psychology*. It is not the version of record and is therefore not suitable for citation.
Copyright: Jansen, A., Melchers, K. G., Kleinmann, M., Lievens, F., Brändli, M., Fraefel, L., & König, C. J. (2013). Situation assessment as an ignored factor in the behavioral consistency paradigm underlying the validity of personnel selection procedures. *Journal of Applied Psychology*, 98, 326–341. doi:10.1037/a0031257

**Situation Assessment as an Ignored Factor in the Behavioral Consistency Paradigm
Underlying the Validity of Personnel Selection Procedures**

Anne Jansen and Klaus G. Melchers

Universität Zürich

Filip Lievens

Ghent University

Martin Kleinmann, Michael Brändli, and Laura Fraefel

Universität Zürich

Cornelius J. König

Universität des Saarlandes

Author Note

Anne Jansen & Klaus G. Melchers, Department of Psychology, Universität Zürich, Switzerland; Filip Lievens, Department of Personnel Management and Work and Organizational Psychology, Ghent University, Belgium; Martin Kleinmann, Michael Brändli, and Laura Fraefel, Department of Psychology, Universität Zürich; Cornelius König, Department of Psychology, Universität des Saarlands, Saarbrücken, Germany.

The first three authors contributed equally to this research. We thank Isabelle Odermatt for her help with the pretests of assessment center exercises and interview questions, Kurt Ackermann for his help with the data collection and data coding, Thomas Staufenbiel for statistical advice, and Ute-Christine Klehe for helpful comments on a previous version of this manuscript.

Correspondence concerning this article should be addressed to Anne Jansen, Department of Psychology, University of Zurich, Binzmühlestrasse 14/12, CH-8050 Zurich, Switzerland. Electronic mail may be sent to a.jansen@psychologie.uzh.ch.

Abstract

This study contributes to the literature on why selection procedures that are based on the behavioral consistency logic (e.g., structured interviews and assessment centers) are valid predictors of job performance. We rely on interactionist theories to propose that individual differences in assessing situational demands explain true variance in performance in selection procedures and on the job. Results from 124 individuals in a simulated selection process showed that the assessment of situational demands was related to both selection and job performance. Individual differences in assessing situational demands also contributed to the criterion-related validity of assessment center and structured interview ratings, offering a complementary explanation as to why selection procedures based on the notion of behavioral consistency predict job performance.

Keywords: behavioral consistency, job performance, employment interview, assessment center, assessment of situational demands

Situation Assessment as an Ignored Factor in the Behavioral Consistency Paradigm Underlying the Validity of Personnel Selection Procedures

Since Wernimont and Campbell's (1968) seminal article, the notion of behavioral consistency has been used as an explanation for the criterion-related validity of selection procedures such as assessment centers (ACs), accomplishment records, structured interviews, work sample tests, or behavioral tendency situational judgment tests (SJTs). Behavioral consistency is typically defined in terms of maximizing the point-to-point correspondence between candidates' behavior in the selection procedure and their future work behavior (e.g., Schmitt & Ostroff, 1986). Although the behavioral consistency logic has become standard terminology in explaining the validity of selection procedures it is also important to understand which factors make people's behavior consistent across the two contexts (i.e., the selection and the work context). Prior research has begun to examine the consistency between candidates' behavior in selection procedures (typically captured in ratings of candidates' knowledge, skills, and abilities, KSAs) and their future work behavior by focusing on the underlying common ability and personality determinants. For instance, KSAs measured in structured interviews and assessment center exercises have been linked to cognitive ability and personality (e.g., Berry, Sackett, & Landers, 2007; Collins et al., 2003; Dilchert & Ones, 2009; Huffcutt, Conway, Roth, & Stone, 2001; Huffcutt, Roth, & McDaniel, 1996; McDaniel, Hartman, Whetzel, & Grubb, 2007; Meriac, Hoffman, Woehr, & Fleisher, 2008; Roth, Bobko, & McFarland, 2005; Salgado & Moscoso, 2002). This stream of research is important because it supports one explanation for the validity of these procedures for predicting job performance. That is, behavior on the job and KSA ratings of candidates' behavior in selection procedures share similar ability and personality determinants.

We posit that the focus in prior research on ability and personality determinants in explaining selection performance and validity represents only one side of the equation. Our

arguments are based on interactionist theories that propose that behavior (within both work and selection contexts) results from an interaction between the person and the perception of the situation (Funder, 2006; Reis, 2008; Tett & Burnett, 2003). Specifically, interactionist theories have emphasized that people's behavioral manifestations are based not only on their standing on traits (personality, ability) but also on how they assess the situation (Mischel & Shoda, 1995; Tett & Burnett, 2003).

Consistent with these recent theories about person-situation contingencies, the present study focuses on candidates' assessment of situational demands. Our basic premise is that the criterion-related validity of selection procedures that are based on the behavioral consistency paradigm is also due to the fact that effective behavior in both selection and job contexts requires assessing the demands of the situations encountered. Up to this point, individual differences in assessing situational demands have been largely ignored as a complementary explanation of people's behavioral consistency across selection and work contexts.

Theoretical Background

From an interactionist perspective, behavior is a function of the interaction between the person and that person's perception of the situation (Endler & Magnusson, 1976; Funder, 2006; Pervin, 1989; Reis, 2008; Tett & Burnett, 2003). For years, researchers have searched for a useful way of classifying situations (Ekehammar, 1974), producing a myriad of situational attributes (for a review of situational taxonomies, see Ten Berge & De Raad, 1999). Therefore, an important development was the distinction between "nominal" and "psychological" situations (Block & Block, 1981). A nominal situation represents the stimulus context as it is generally understood (Block & Block, 1981; Reis, 2008). For a situation to be nominal (also called consensual), many individuals rate the demand qualities of the situation and agree on them. In contrast, the psychological situation (also known as the functional situation) represents the situation as it is perceived by a particular individual.

The importance of a particular individual's assessment of the situation has been well articulated in the Cognitive-Affective Personality System (CAPS) theory (Mischel & Shoda, 1995, 1998). This comprehensive interactionist theory posits that nominal features of situations activate a series of mental representations (both cognitive and affective). On the basis of these particular interconnected and interacting CAPS units being activated, specific behavioral scripts are triggered. One key implication of the CAPS theory for this study is that it provides one way of understanding the linkages between situations as they are nominally perceived, and situations as they are perceived by individuals. It is this psychological situation that is the focus of the present study. In particular, this study focuses on one key cognitive unit of the CAPS theory. Specifically, we measure how individuals assess situational demands and relate this assessment to their selection and job performance.

Another key implication of the CAPS model is that it posits that people differ in their assessment of situational demands. As Mischel and Shoda (1995) noted: "Individuals differ in how they selectively focus on different features of situations, how they categorize and encode them cognitively and emotionally, and how those encodings activate and interact with other cognitions and affects" (p. 252). Since the development of the CAPS theory, research has confirmed that individuals reliably differ in how they encode and process situations (see also Fleeson, 2007). In addition, such differences in how people assess the demands of a given situation seem to lead to variability across people in their subsequent behavioral manifestations in that situation.

Using CAPS theory as a general framework, the present study puts forth the general hypothesis that individual differences in assessing situational demands will: (a) relate to selection and job performance; and (b) serve as an untested explanation of people's consistency across these two contexts (i.e., selection and work). As described below, previous research has shown that individual differences in assessing situational demands encountered

in selection procedures that are based on the behavioral consistency paradigm are positively related to performance in these procedures. However, at this point, we do not know whether differences in assessing situational demands represent an ignored “third variable” in explaining the validity of these procedures (apart from the KSAs typically assessed in them).

The remainder of the Introduction develops this general hypothesis by elucidating the role of assessing situational demands in job performance and personnel selection. We will exemplify it using ACs and interviews. Before doing so, however, we first define the construct of the assessment of situational demands, discuss to what degree it is related to or distinct from other constructs, and review prior research.

Assessing Situational Demands: Definition and Related Constructs

Assessing the demands of situations refers to cognitive processes for deciphering (or “reading”) what is required to behave effectively in situations. In organizational situations, what is considered to be effective behavior is typically reflected in “performance criteria” (see Binning & Barrett, 1989; Schleicher & Day, 1998). Therefore, within work and organizational contexts, assessing situational demands refers to individuals’ cognitive processes to decipher the performance criteria in evaluative situations.

In line with CAPS theory, individuals differ in their assessment of situational demands. People who are good at the perceptual skill of situation assessment engage in cognitive processes to find out which criteria are being used to judge their performance in evaluative situations. To this end, they scan the situation to pick up cues. Furthermore, they distinguish between relevant and irrelevant cues, see relationships among the multiple cues, validate prior cues with newer ones, and use this information to infer the key situational demands. By doing so, people who are good at situation assessment also “read” the situational demands in line with the performance criteria as consensually defined in a particular organization. Given that the aforementioned mental activities inherent in situation

assessment are cognitively complex, people who are higher on cognitive ability should have an advantage in engaging in these information processing activities to decipher the situational demands. As described below, previous research has generally confirmed this link between cognitive ability and situation assessment.

The essence of the assessment of situational demands is further clarified by contrasting this construct with social effectiveness constructs (known under various aliases such as social skill, social competence, social intelligence, political skill, etc.). According to Ferris, Perrewé, and Douglas (2002) social effectiveness refers to “the ability to effectively read, understand, and control social interactions” (p. 49). Social effectiveness is a broader multidimensional term that encompasses the cognitive element of reading and understanding social situations as well as the behavioral component of being able to act on that insight to influence others (Ferris et al., 2002). So, social effectiveness includes “the dual components of understanding people and social situations, and being able to act on that knowledge or understanding in an appropriate manner” (Ferris et al., 2002, p. 50).

Situation assessment and the social effectiveness constructs share the cognitive dimension of reading situational demands. However, even though situation assessment is obviously an important basis for people’s subsequent behavior, it does not include these subsequent behavioral adaptations. This constitutes a difference with social effectiveness constructs, which include both a cognitive as well as a behavioral component. Another difference between situation assessment and the social effectiveness constructs is that situation assessment is restricted to evaluative situations and deals with deciphering the performance criteria.

Given these conceptual distinctions between the assessment of situational demands and social effectiveness constructs, the measurement of these constructs also differs. Whereas social effectiveness constructs are usually conceptualized in a trait-like manner and measured

via self-reports (e.g., Ferris et al., 2002), the result of the cognitive processes underlying the assessment of situational demands is measured with a test in which there are correct and incorrect answers. Essentially, evaluating individuals' assessment of situational demands is based on comparing their particular perceptions of the performance criteria (i.e., their assumptions regarding what is being measured) in a given evaluative situation to the consensually-defined (i.e., nominal) performance criteria.

On the basis of previous research by Kleinmann (1993), König, Melchers, Kleinmann, Richter, and Klehe (2007) developed such a measure of people's assessment of situational demands (see also Jansen, Lievens, & Kleinmann, 2011). This was an important methodological advancement because it sidestepped the biases in self-reports. In this measure, individuals are asked to write down the demand qualities on which they think they were being evaluated during evaluative situations (e.g., selection procedures). People's answers are then matched to the consensually-defined situational demand qualities (the performance criteria used by the organization in the selection process). König et al. labeled this measure "ATIC", which stands for the Ability to Identify Criteria.

Assessing Situational Demands: Prior Research

Prior empirical research supports the aforementioned theoretical propositions underlying the assessment of situational demands by placing the construct in a nomological network. Table 1 summarizes this prior research. First, Table 1 shows that moderate but consistent correlations between cognitive ability and the assessment of situational demands were found. Thus, the link between cognitive ability and situation assessment is well established. Second, as social effectiveness constructs also include a cognitive component (albeit measured through self-reports), it comes as no surprise that significant correlations between many social effectiveness constructs and the assessment of situational demands were found, with correlations between .20 and .25. Third, prior research has revealed that the

relationship between assessment of situational demands and personality traits is inconsistent. Some of these correlations reached absolute values above .30, whereas others were close to 0, a finding that might hint at the existence of moderators. Fourth, prior research also suggests that the ability to assess performance criteria is not entirely invariant of the criteria being assessed (cf. Table B1 in Appendix B, which summarizes all prior studies about situation assessment for which we managed to obtain dimension-specific scores for the situation assessment measure employed). However, it seems likely that there is an interaction between the ability to decipher the criteria being assessed and the specific selection situation. That is, the same criterion (e.g., communication) might be easier to pick up in one situation (e.g., an oral presentation exercise or an interview) than in another one (e.g., an in-basket). Finally, prior research confirmed that there exists consistency in participants' assessment of situational demands across evaluative situations because participants who were good at assessing situational demands in interviews were also good at this in ACs and vice versa (König et al., 2007).

Assessing Situational Demands and Job Performance

To this point, we still do not know whether the assessment of situational demands contributes to performance on the job. We propose that the assessment of situational demands might be relevant in the context of job performance, provided that three broad conditions are satisfied. In line with the CAPS framework, as a first condition, there has to be a trigger to activate the cognitive processes underlying the assessment of situational demands (and the subsequent behavior). As mentioned in the definition of assessment of situational demands, evaluative situations in organizations are considered such activators. Hereby we stress that it is important that people know that they are being evaluated so that they will actively start searching for cues about the criteria being used to evaluate their behavior. Along these lines, Morrison (1993) gave an overview of the different ways of gathering information regarding

performance criteria in the workplace. These approaches include monitoring supervisors/co-workers, consulting written material (e.g., mission statement of the organization, performance appraisal forms), and directly asking supervisors/co-workers about the performance criteria.

As a second condition, relevant cues about situational demands have to be available and it must be necessary for individuals to pick up and process these cues in order to successfully decipher these demands. Conversely, if the demands of the situation (performance criteria) are very clear and known to everyone (e.g., in piece work), then there is neither a need to search for cues nor a need to assess the demands of the situation. So, this condition refers to the clarity of the evaluative situation, which denotes the extent to which cues regarding situational demands are available and easy to decipher (Meyer, Dalal, & Hermida, 2010). This condition is important because it ensures that engaging in cognitive processes to decipher situational demands has effects, and that individual differences in assessing situational demands emerge.

Third, as an outcome of this process, individuals are expected to demonstrate behaviors in the evaluative situation that depend on their interpretation of that situation. Thus, consistent with the CAPS theory, the behaviors shown will not only be based on individuals' KSAs but also on their assessment of the situational demands. As such, the assessment of situational demands serves as a compass to decide which specific KSAs to use (Hogan & Shelton, 1998). Specifically, employees need to assess the situational demands they encounter and exhibit behavior that fits these situational demands (Tett & Burnett, 2003). If their interpretation of the situational demands is in line with the consensually-defined situation, they will be able to exhibit more effective behavior and therefore obtain higher performance ratings. Conversely, misinterpretation of the performance criteria might lead them to display behaviors that are not demanded by the situation or that even run counter to the situational demands, resulting in low performance ratings. With respect to the third

condition, it is noteworthy that the assessment of situational demands constitutes an implicit element of common job performance definitions. A common thread running through job performance definitions is that for behavior to reflect effective performance, this behavior needs to be evaluated with respect to how it is congruent with a broader context (see Campbell, 1990; Motowidlo, Borman, & Schmit, 1997; Viswesvaran, 2000).

These three conditions are present in many organizational situations. Straightforward examples are starting to work in a new job (the probationary period) or on a new task assignment. More generally, however, we believe that the assessment of situational demands will be particularly important for showing effective behavior in jobs wherein employees are confronted with novel or unpredictable situations and make decisions about unfamiliar topics (see Murphy's 1989, notion of transitional job stages).

When these three conditions are present, we expect that the assessment of situational demands will represent substantive legitimate variance instead of unwanted error variance (e.g., contamination, deficiency) in job performance. Thus, we expect a link between the assessment of situational demands and job performance and therefore we formulate the following hypothesis:

Hypothesis 1: Individual differences in assessing situational demands will predict job performance.

Assessing Situational Demands and the Behavioral Consistency Paradigm in Selection

Above we posited that an adequate assessment of situational demands contributes to employees showing effective behavior in evaluative work situations within organizations. Against the backdrop of the three aforementioned conditions, it becomes clear that the assessment of situational demands has important implications for personnel selection procedures that are based on the behavioral consistency paradigm. First, the selection situation is a situation where applicants know that they are being evaluated. Hence, the

selection situation will typically trigger applicants to find out what is actually being assessed. Research has indeed confirmed that applicants actively search for cues about the situational demands (i.e., the performance criteria being used to evaluate them) in selection (Alexander & Knight, 1971; Alexander & Rudd, 1984; Tullar, 1989).

The second condition is met because it is typically not clear to applicants what is being evaluated in the general selection process and in the specific selection procedures (Tullar, 1989). In interviews, for example, candidates are often not informed about the particular requirements tied to interview questions (Ferris & Judge, 1991; Fox & Spector, 2000; Latham & Sue-Chan, 1999). This is definitely the case in unstructured interviews, but may also be true of structured interviews. For example, situational questions usually contain a situational dilemma (i.e., the actions to be taken are equally [un]desirable, cf. Latham & Sue-Chan, 1999; Maurer, Sue-Chan, & Latham, 1999; Motowidlo, 1999). Similarly, in AC exercises it is often unclear to candidates which performance criteria are assessed (Kleinmann, 1993; McFarland, Yun, Harold, Viera, & Moore, 2005).

So, to decipher the performance criteria applicants need to detect and process relevant cues. For example, regarding the selection process, applicants might scrutinize the job ad and/or browse the company's website. In addition, experiences in prior selection procedures and (in)formal coaching might be informative to find out situational demands. Specific selection procedures might also provide cues about what is being assessed. For instance, in an interview, candidates might attempt to read between the lines of interviewers' questions or look for nonverbal cues (e.g., when do interviewers note something down?). As another example, in an AC exercise, applicants' assessment of the situational demands might be influenced by reading the instructions or observing other participants' behavior. As it is often not clear what is being assessed in the general selection process and in specific selection

procedures, applicants might differ from each other in how they assess the situational demands.

As for the third condition, many selection procedures (e.g., ACs, work samples, accomplishment records, behavioral tendency situational judgment tests, interviews) require a behavioral response associated with the performance criteria deciphered, thereby invoking the behavioral consistency paradigm to enable predictions about future candidate behavior. In particular, we posit that, apart from the KSAs that applicants possess, their interpretation of the situational demands might enable them to show effective behavior. Conversely, misinterpretation of the situational demands on their part might lead them to display behaviors that run counter to the consensually-defined performance criteria, resulting in low selection ratings. As such, individual differences in situation assessment will explain variance in selection performance.

Generally, prior research supports these assumptions. For instance, Kleinmann (1993) showed that candidates differed considerably in the number of performance criteria (ranging from 0 to 16 out of 20 across all 5 exercises) correctly identified in an AC, with most of the participants correctly identifying only 5 to 8 criteria. Furthermore, as shown in Table 1, previous studies found the assessment of situational demands to be positively related to interview (r s from .23 to .35) and AC performance (r s from .23 to .49). This was even true when the assessment of situational demands was measured in one selection procedure to predict performance in another selection procedure (König et al., 2007). Thus, candidates who were good at deciphering the situational demands in selection procedures performed better in these procedures. Accordingly, we also expect a link between the assessment of situational demands and selection performance in the present study and therefore formulate the following hypothesis:

Hypothesis 2: Individual differences in assessing situational demands will predict performance in behavior-based selection procedures.

Our two stated hypotheses posit significant correlations between situation assessment and selection and job performance, respectively. As evaluations of the criterion-related validity of selection procedures deal with estimating relationships between predictor (selection) and criterion (performance) scores, it follows from Hypothesis 1 and Hypothesis 2 that situation assessment also plays a key role in the criterion-related validity of selection procedures that are based on the behavioral consistency logic (e.g., ACs and structured interviews). We posit that situation assessment will explain some of the common variance between selection performance and job performance. Specifically, we argue that situation assessment serves as a compass that generally guides people's behavior in the direction demanded by the situation. Accordingly, we assume that individual differences in people's ability to assess situational demands contribute to both their performance in behavior-based selection procedures as well as to their performance on the job. Thus, variance in selection performance that is due to these individual differences should be related to variance in job performance that is due to the same individual differences. Or said differently, individual differences in people's ability to assess situational demands are a common source of variance for performance in both contexts.

When one statistically models this common source in a model that includes: situational assessment, performance in behavior-based selection procedures, and job performance; then the direct path between selection performance and job performance should be weaker than the same path in a rival model not including a direct path from assessment of situational demands to job performance. This leads to the following hypothesis:

Hypothesis 3: The relationship between performance on behavior-based selection procedures and job performance will significantly decrease when controlling for individual differences in assessing situational demands.

Method

Participants

Participants were invited via e-mail to take part in a simulated selection process. Participation in this simulation would enable them to gain experience in various selection procedures and to receive feedback on their selection performance. In this study, we tested our hypotheses in a sample wherein participants had just started a job or had only limited work experience. So, we kept two main considerations in mind when including participants in the sample. One consideration was that all participants had to have previous work experience during the six-month period prior to participating in the simulated selection process. As a minimum, they needed to have worked more than 12 hours per week. Study participation was voluntary on the condition that participants provided their recent or current supervisor's e-mail address and permission to have this supervisor assess their job performance. At the same time, as another consideration, we focused on individuals with only limited work experience (e.g., newcomers in a job). There were no exclusion criteria with regard to the type of job. Most of the jobs included were internship jobs, part-time jobs, or temporary jobs in various domains. According to information provided by the supervisors, 56.1% of the jobs were in the educational and research sector, 16.8 % in the service sector, 9.3% in the industry, 6.5% in banking and insurance, and 11.3% in other sectors.

In light of these considerations, our sample ($N = 124$; 67 males, 57 females) consisted of either prospective or recent graduates from several Swiss universities. Participants' mean age was 27.6 years ($SD = 4.17$). More than half of them (55.3%) held a Master's degree. Of these, 35.3% were pursuing a PhD and about 6.5% already held a PhD. Participants came

from various majors, with most of them having a background in sciences (40.3%). The others majored in humanities (16.1%), business or economics (15.3%), engineering (12.1%), law (4.0%), or other subjects (8.1%). Four percent did not indicate their major.

Several efforts were undertaken to make sure participants were motivated and behaved like “real” applicants. One such effort was that participants self-selected to participate as they had to register on a website. To further ensure that only highly motivated people participated, people had to pay a small fee to cover part of the costs. In addition, prizes were awarded: A cash prize equivalent to US \$100 was given to the best candidate and about US \$50 to the second best candidate per day. Finally, assessors and most of the participants were formally dressed.

Simulated Selection Process

The simulated selection process consisted of four AC exercises, a structured interview, and a cognitive ability test. Given the well-established relationship between cognitive ability and situation assessment, the test was included to take cognitive ability into account as an antecedent of situation assessment, and to avoid potential concerns that the impact of situation assessment might be overestimated when cognitive ability is not taken into account. All selection procedures were designed to simulate a selection process for a graduate trainee position because such a position represents a realistic and attractive job for applicants with various study backgrounds. A maximum of 12 participants took part in each selection simulation, which lasted for a whole day. First, participants received a job ad that described the key job requirements and watched a short presentation on the selection process. Next, participants completed the AC exercises (see Appendix A for a description of these exercises), the structured interviews, and the cognitive ability test in varying order. In addition, participants had to complete several questionnaires unrelated to the present study.

After each AC exercise, participants completed a situational assessment questionnaire (see below).

In each exercise, two assessors observed and evaluated the performance of each participant. The assessors also conducted the structured interviews, with two assessors interviewing each participant. One assessor read the questions to the participant, and both assessors independently scored the responses on the basis of the scoring guide before asking the next question. In total, seven dimensions were assessed in the selection procedure: organizing, consideration of others, persuasiveness, analytical skills, presentation skills, assertiveness, and creativity. The maximum number of dimensions per AC exercise was four. Interview questions were developed to assess either consideration of others or organizing.

Assessors came from a pool of 31 Master's-level students (25 females, average age 25.8 years, $SD = 3.84$), most of whom were pursuing their Masters' degree in industrial and organizational (I/O) psychology. They had been trained in a one-day training session. In the training, they were introduced to the exercises, the dimensions, and the rating sheets for the AC exercises, which gave behavioral examples of the dimensions. They were also introduced to the interview questions and the interview scoring guide. They took turns participating in the exercises themselves and practicing the observation and evaluation of performance in the exercises. To achieve a consistent frame-of-reference (Woehr & Huffcutt, 1994), they discussed their ratings afterwards. Additionally, they learned about typical rating errors and received information about how to conduct the feedback session.

At the end of the selection simulation, participants had to answer several demographic questions as well as questions concerning the perceived realism of the simulation. Afterwards, they received dimension-level feedback on their performance in the AC exercises and the interview. Finally, the two top participants received their cash prizes.

Taken together, the above description shows that we did our best to guarantee the external validity of the selection simulation. The simulation had several elements that are similar to actual selection practice (Eurich, Krause, Cigularov, & Thornton, 2009; Spychalski, Quiñones, Gaugler, & Pohley, 1997) and in line with the guidelines and ethical considerations for AC operations (International Task Force on Assessment Center Guidelines, 2009). Specifically, (a) participants received contextual information (i.e., a job ad that described the requirements of the job) prior to participating, (b) a broad variety of tests, interviews, and AC exercises representative of actual selection practices were used, (c) the exercises and interview questions were designed to evoke behaviors for evaluating relevant dimensions, (d) trained assessors with an I/O background served as assessors, (d) multiple assessors evaluated each participant, and (e) participants received detailed feedback about their performance afterwards. When asked about the perceived realism of the selection process in a post-administration survey, 85.4% of the participants stated that they had acted like real applicants.

Measures

AC performance. For each exercise, assessors rated participants' performance on each targeted dimension on a 5-point scale (ranging from 1 = poor performance to 5 = excellent performance). After participants had completed all exercises and the interview, assessors discussed their ratings when they differed by two or more points and adjusted their ratings accordingly. Following that, the ratings of the two assessors were averaged. To assess inter-rater reliability, we calculated intraclass correlations (ICC 1,1) between the dimensional ratings of the two assessors. The mean inter-rater reliability (i.e., the reliability of a single assessor) was .81. For our analyses, we calculated overall dimension ratings (across all exercises per dimension) and an overall AC rating (across all exercises and all dimensions).

Interview performance. Interview performance was rated on a 5-point scale.

Behavioral anchors were provided for 1 = *poor performance*, 3 = *average performance* and 5 = *excellent performance*. As was the case for AC performance, interviewers discussed their ratings when they differed by two or more points, and the agreed upon ratings of the two interviewers were averaged. To assess inter-rater reliability, we calculated intraclass correlations (ICC 1,1) between the ratings of the two interviewers. The mean inter-rater reliability (i.e., the reliability of a single interviewer) was .82. For our analyses, we calculated overall dimension ratings (across all questions per dimension) and an overall interview rating (across all questions and all dimensions).

Cognitive ability. We used a German translation of the Wonderlic Personnel Test (Wonderlic Inc., 2002) to measure cognitive ability. This 12-minute, 50-item test assesses vocabulary, arithmetic reasoning, and spatial relations and is a commonly used measure of cognitive ability. In our study, the coefficient alpha was .82.

Candidates' assessment of situational demands. In measuring candidates' assessment of situational demands we focused only on the AC exercises for two reasons. First, AC exercises are work samples that aim to provide a much closer approximation of the actual work context than interviews. The idea was that if candidates are able to assess situational demands in AC exercises, they might also be able to do so in the work context (behavioral consistency logic). And second, as noted above, prior research found that assessment of situational demands in one selection procedure (e.g., an AC) predicts performance in other selection procedures (e.g., an interview) when all these selection procedures are part of the same selection process (König et al., 2007). Therefore, we also used the situational demands score (obtained via participation in AC exercises) in our analyses with the interview.

Determining candidates' assessment of situational demands involves measuring both the nominal situation and the psychological situation. The following describes how we measured both of those situations.

Nominal situation. Nominal situations refer to situational attributes that are defined independently from any person (e.g., by aggregating the situational demands across a number of experienced observers, Block & Block, 1981; Reis, 2008). To determine the dimensions targeted in the selection procedure, the AC exercises (instructions, descriptions) were provided to a group of 13 I/O psychology graduate students. Those students were knowledgeable about human resource management and ACs. They inspected the documents and rated how well the exercises evoked the targeted dimensions in contrast to various other dimensions. The dimensions that were rated as the most likely ones to be assessed in the AC exercises were considered to reflect the demands of the nominal situation. Those dimensions served as the benchmark to which we could compare each candidate's particular perception of the situational demands (see below).

Psychological situation. In line with König et al. (2007), participants had to answer a questionnaire after each exercise that was designed to assess participants' perceptions of the demands of the situation in the AC exercises. Specifically, they were instructed to write down their assumptions about what dimensions were being assessed in the exercise. In addition to the procedure used by König et al. (2007), participants were asked to give behavioral examples for their assumptions. To ensure that participants understood this procedure, they received an example. They were encouraged to write down as many dimensions and behavioral examples per exercise as they thought of during the exercise.

Measure of candidates' situation assessment. Two pairs of external raters (a trained Master's-level student in I/O psychology and three authors of the present study) rated the degree to which each of the participants' assumptions and behavioral examples

(psychological situation) corresponded to the targeted dimensions (nominal situation). The raters examined the assumptions and the behavioral examples and rated the fit with the targeted dimensions on a scale from 0 = *no fit* to 3 = *fits completely*. If none of the assumptions was related to a targeted dimension, a score of 0 was assigned. In case of ties (several assumptions being linked to the same dimension), we used the highest strength of fit rating as the score. After coding the assumptions, the raters discussed their ratings when the ratings differed by two or more points (on the 4-point scale). The reader is referred to Jansen, Lievens, and Kleinmann (2011) for a detailed description of the measure. To determine inter-rater reliability, we calculated intraclass correlations (ICC 1,1) between two raters. The average inter-rater reliability was .86.

To calculate scores of candidates' assessment of situational demands, which could range between 0 (= *poor assessment of situational demands*) and 3 (= *excellent assessment of situational demands*), the ratings were averaged across exercises to obtain dimension-specific scores, and across dimensions and exercises to obtain an overall score. We decided to focus on the overall score for our later analyses because of its higher internal consistency reliability. Given that each dimension was rated in no more than four exercises (with two dimensions that were rated in only one exercise) the internal consistency reliability of dimension specific scores for both participants' assessment of situational demands and the corresponding performance ratings was low.

Job performance. Participants' supervisors were sent the link to an online questionnaire and were asked to evaluate participants' job performance. Of the supervisors contacted, 107 (86.3%) answered the questionnaire. More than half of the supervisors had held their position for more than six years (53.3%), 23.4% had held such a position for more than three years, and 23.3% had less than three years of experience as a supervisor. Supervisors did not have access to participants' ratings in the selection procedures.

Job performance was measured on a 7-point scale ranging from 1 = *not at all* to 7 = *absolutely*, with five items from Williams and Anderson (1991) in their German translation by Staufenbiel and Hartz (2000) that assess in-role behavior, and five items from the task-based job performance questionnaire by Bott, Svyantek, Goodman, and Bernal (2003) translated by the authors of this article. For our analyses, we computed the average across all items because of the high intercorrelation among the two scales ($r = .79$). Coefficient alpha of this composite scale was .92.

Results

Preliminary Analyses

Table 2 presents means, standard deviations, and correlations of the study variables. In one case, age was missing and was replaced by the sample's mean. The mean score of cognitive ability was 31.71 and the standard deviation was 5.72, which means that our sample had higher scores and a smaller standard deviation than the population norm (Wonderlic Inc., 2002) for college graduates ($M = 29.60$, $SD = 6.30$) or the adult working population in general ($M = 21.75$, $SD = 7.60$). The mean score on the overall measure of assessment of situational demands was $M = 1.54$ on a scale from 0 to 3, indicating that it was not easy for participants to identify the targeted dimensions in the AC exercises. The standard deviation of this measure was .40, which indicated that participants differed in the degree to which they identified the dimensions targeted (see Kleinmann, 1993). Furthermore, across the different dimensions, the mean scores for participants' assessment of situational demands varied between .80 (for analytical skills) and 2.21 (for presentation skills), which suggests that some of the dimensions were easier to decipher in the context of the present AC exercises than others.

Both the AC and interview had adequate criterion-related validity. The overall AC rating, which was determined as the average of the dimension ratings across exercises, was

significantly correlated with job performance, $r = .21, p < .05$, with the presentation skills dimension emerging as the strongest predictor, $r = .25, p < .05$. Similarly, the overall interview rating, which was determined as the average of the dimension ratings across interview questions, was also significantly correlated with job performance, $r = .19, p < .05$. In the interview, the organizing dimension was the strongest predictor, $r = .19, p = .05$. Regarding the relation between cognitive ability and AC performance, we found a significant correlation, $r = .27, p < .05$, which is in line with the results of previous meta-analyses (Collins et al., 2003; Hoeft & Schuler, 2001). The correlation between cognitive ability and interview performance was also significant, $r = .22, p < .05$, thereby also confirming the findings of previous meta-analyses (Berry et al., 2007; Huffcutt et al., 1996). The relation between cognitive ability and job performance was $r = .18, ns$, which is somewhat lower than the meta-analytic correlations usually reported in the literature but still within the 90 percent confidence interval of uncorrected meta-analytic estimates (Salgado, Anderson, Moscoso, Bertua, & de Fruyt, 2003).

Finally, in line with prior research on the assessment of situational demands, there was a moderate correlation between cognitive ability and candidates' assessment of situational demands, $r = .29, p < .05$.

Test of Hypotheses

According to Hypothesis 1, individual differences in assessing situational demands will predict job performance. In line with this, we found that candidates' assessment of situational demands was significantly related to job performance, $r = .27, p < .05$ meaning that people who were better at assessing the situational demands in the selection procedure received higher performance ratings from their supervisors.

Next, Hypothesis 2 predicted a correlation between participants' assessment of situational demands and their selection performance. In line with this prediction, there were

significant correlations between situation assessment and the overall interview rating, $r = .26$, $p < .05$, and the overall AC rating, $r = .23$, $p < .05$, respectively. Hence, participants who were good at assessing the situational demands received better performance ratings in the interview and the AC exercises.

Furthermore, Hypothesis 3 posited that the correlation between interview performance and job performance, and between AC performance and job performance, respectively, will decrease when controlling for candidates' assessment of situational demands. To test this, we calculated the correlation between overall performance ratings in the two selection procedures and job performance by partialling out candidates' assessment of situational demands. The correlation between job performance and the overall interview rating dropped to $r = .13$, *ns*, compared to the prior zero-order correlation of $r = .19$, $p < .05$. Statistically controlling for the impact of individual differences in assessing situational demands in the relation between job performance and overall AC rating lowered the correlation to $r = .17$, *ns*, compared to the zero-order correlation of $r = .21$, $p < .05$. This means that the relations between performance in the selection procedures and job performance were no longer significant when participants' assessment of situational demands was controlled for.

Test of Hypothesized Model

We used structural equation modeling (via Amos 18) to test a model including all the links proposed in the hypotheses as well as cognitive ability (as a well-established antecedent of situation assessment). For the model test, both cognitive ability and job performance were defined by three parcels of items. Starting from Items 1, 2, and 3, parcels were built by taking every third item and determining the average value of these items. Candidates' assessment of situational demands was defined as the average of the fit ratings for the targeted dimensions for two parcels of items. Specifically, we did an odd-even split across all AC exercises and dimensions and calculated to what degree participants had identified the corresponding

dimensions. Finally, we specified AC performance by using the seven overall dimension ratings as indicators, and we modeled interview performance with the two overall dimension ratings of the interview as indicators. We used maximum likelihood estimation and tested this model once with the AC as the selection procedure and once with the interview as the selection procedure.

Concerning the model including the AC, overall goodness-of-fit indices showed that the hypothesized model produced a good fit to the data, $\chi^2(86) = 112.67, p < .05$, TLI = .96, IFI = .96, CFI = .96, and RMSEA = .05 [90% CI = .02 - .08]. Parameter level estimates are shown in the top panel in Figure 1. Results of the measurement model showed that each of the latent factors was reliably measured. Concerning the structural model, several important findings emerged. First, the path coefficient from cognitive ability to situation assessment was significant, indicating that cognitive ability predicts assessment of situational demands. Second, the path coefficients from assessment of situational demands to both AC performance and job performance were significant, indicating that people who were better at assessing the situational demands not only received higher dimensional ratings in the AC, but also performed better on the job. Third, the path from AC performance to job performance was not significant. This is in line with the argument that individual differences in assessing situational demands contribute to the positive relationship between AC performance and job performance and that this relationship decreases when individual differences in assessing situational demands are taken into account as a common cause.

As an additional test of Hypothesis 3 and to further evaluate the appropriateness of this model, we tested an additional model that did not include a direct path from assessment of situational demands to job performance. This model also had a very good fit, $\chi^2(87) = 120.73, p < .05$, TLI = .94, IFI = .94, CFI = .95, and RMSEA = .06 [90% CI = .03 - .09]. Parameter level estimates for this model are presented in the bottom panel in Figure 1. As

shown in the figure, the path from AC performance to job performance was significant in this model, meaning that AC performance is a significant predictor of job performance when assessment of situational demands is not modeled as a common cause of both of these factors. Furthermore, in this case the impact of individual differences in assessing situational demands on job performance is mediated by AC performance. A comparison of both models with a χ^2 -difference test revealed that the common cause model fit the data significantly better, $\chi^2(1) = 8.06, p < .05$, meaning that the common cause model is more appropriate.

Testing these two models for the interview yielded virtually the same findings and conclusions. Specifically, both models produced a good fit to the data and results of the measurement model showed that each of the latent factors was reliably measured. As with the AC, the path from interview performance to job performance was not significant for the model including a direct path from assessment of situational demands to job performance but became significant in the second model that did not include a direct path from assessment of situational demands to job performance. Furthermore, a comparison of both models by means of a χ^2 -difference test again revealed that the common cause model fit the data significantly better, $\Delta\chi^2(1) = 7.17, p < .05$, indicating that the common cause model is more appropriate. Detailed results of these analyses can be obtained from the authors.

Discussion

Main Contributions

In personnel selection, there has been a longstanding interest in constructs underlying the validity of selection procedures that are based on the behavioral consistency paradigm such as structured interviews and ACs. Most attention has been paid to ability and personality as underlying determinants of the KSAs measured in these selection procedures. The present study relied on interactionist theories (Fleeson, 2007; Mischel & Shoda, 1995; Reis, 2008;

Tett & Burnett, 2003) to posit that selection and job performance are also dependent on how candidates assess situational demands.

To the best of our knowledge, our study is the first to test the key question of whether individual differences in assessing situational demands represent “noise” or substantive legitimate variance in selection and job performance. Our findings revealed that individuals who were better able to assess the situational demands in the selection procedure not only showed better selection performance but also received higher job performance ratings by their supervisors. In addition, the correlations between candidates’ selection ratings and job performance were no longer significant when controlling for their assessment of situational demands in those selection procedures. In short, these results provide evidence that individual differences in assessing situational demands serve as a complementary factor that explains true variance in selection and job performance. This adds a different angle to our understanding of the behavioral consistency paradigm and of the validity of behavior-based selection procedures.

Apart from shedding light on the behavioral consistency paradigm, the result that candidates’ assessment of situational demands represents substantive variance also has important implications for at least two other streams of research. First, it has key implications for our understanding of the validity of structured interviews and ACs. Apparently, not only ability and personality are related to performance in selection situations and on the job. Individual differences in assessing situations are also relevant. To put it another way, in evaluative situations where people have to pick up cues to decipher the situational demands (performance criteria), they will often differ in how well they assess these situational demands, and thus subsequently show different behaviors according to the demands assessed. Apparently, those individual differences in people’s assessment of situational demands matter as they help to explain why structured interviews and ACs predict job performance.

It is noteworthy that candidates' scores on the assessment of situational demands in the AC were used in the interview analyses, predicted interview performance, and contributed to the interview's validity for predicting job performance. As explained above, an individual's score on the situation assessment construct denotes that person's standing on his/her ability to decipher the performance criteria in evaluative situations. Therefore, people's situation assessment can be measured in one situation (e.g., an AC) and can be used as an indication of their ability to decipher performance criteria in analyses related to their performance in other evaluative situations. Performance on the job is one such situation and interview performance constitutes another. However, it seems unlikely that situation assessment as measured in an AC is directly interchangeable with situation assessment as measured in an interview because the specific situations encountered in a set of AC exercises likely differ from the situation encountered in an interview. Nevertheless, recent evidence has shown that participants who were good at assessing situational demands in an AC exercise were also good at this in other AC exercises even when these other exercises had very dissimilar demands (Speer, Christiansen, Melchers, König, & Kleinmann, 2012). This speaks to the cross-situational consistency of situation assessment and suggests that it reflects a general ability that underlies situation assessment in different situations.

The second domain for which the finding that situation assessment represents substantive variance has implications is the literature on social desirability and impression management. Being able to identify the situational requirements seems to be a prerequisite for impression management (see Levashina & Campion, 2006), as people can only successfully manage their impressions when they succeed in deciphering the performance criteria. Hence, our findings might explain why social desirability and impression management in a selection procedure do not seem to lower the validity of selection procedures (Barrick & Mount, 1996; Ellingson, Smith, & Sackett, 2001; Hough, Eaton,

Dunnette, Kamp, & McCloy, 1990; Ones, Viswesvaran, & Reiss, 1996; Schmitt & Oswald, 2006) and may be related to job performance and performance in other evaluative situations (Biderman, Nguyen, Mullins, & Luna, 2008, April; Blickle, Momm, Schneider, Gansen, & Kramer, 2009; Kleinmann & Klehe, 2011).

Limitations

This study is not without limitations. A first limitation relates to the fact that the participants were not real applicants and that the data were obtained in a selection simulation. However, many participants were currently looking for jobs and used the simulation to prepare for real selection situations. Additionally, most participants stated that they perceived the simulation as fairly realistic and that they acted as they would in an actual selection situation. The use of a sample of university graduates who self-selected to participate resulted in less variance on the cognitive ability test than in the population, providing a possible explanation for the somewhat lower correlation between cognitive ability and job performance. However, it should be noted that other results obtained with the cognitive ability test were consistent with prior research (e.g., the correlation of cognitive ability with interview and AC performance).

Second, some generalizability issues should be noted. This study tested the hypotheses with structured interviews and AC exercises as examples of selection procedures that are based on the behavioral consistency paradigm. Future research is needed to extend our findings with other behavioral consistency selection procedures such as situational judgment tests, accomplishment records, or work samples. Similarly, it is important to examine the relevance of the situation assessment construct in other samples and settings. Hereby we acknowledge that this study's sample inclusion criteria (that were mainly based on tenure) represent were an initial attempt to align the sample with the theoretical criterion space (i.e., the aforementioned three conditions). Future studies are needed to test in which

specific jobs and job settings these three conditions apply. To this end, future studies might use O*NET ratings (e.g., high ratings on ambiguity and low ratings on routinization as proxies for the second theoretical condition, Dierdorff & Morgeson, 2007; Lievens, Sanchez, Bartram, & Brown, 2010) to identify jobs where situation assessment might be especially relevant. We also believe that the assessment of situational demands might be especially important for performance in transitional job stages as compared to maintenance job stages (Murphy, 1989).

Implications for Future Research

Future research needs to investigate how people assess situational demands. It is unclear which cues people rely on. On the job, the three information sources (supervisors, co-workers, and written materials) distinguished by Morrison (1993) provide a useful starting point. In selection, applicants might use information that they previously gathered about the company and the job by inspecting the job ad or the company's website. They might also try to infer what is expected in the situation from exercise instructions and interview questions or from nonverbal behavior of interviewers or other participants. Along these lines, future research might also explore how candidates' assessment of situational demands activates behavior. In line with trait activation theory (Tett & Burnett, 2003), a moderator model might be expected. That is, candidates high on a personality trait (e.g., conscientiousness) who correctly pick up on demands related to that trait (e.g., the assessment that planning and organizing is evaluated) should be more likely to activate behavior in line with the cues gathered, leading to better performance. Conversely, without perceiving the situational demands, people high on the related trait might not be able to express behavior related to that trait and thus might show lower performance (see Jansen, Lievens, & Kleinmann, 2011). Further research should therefore investigate whether differences in candidates' assessment

of situational demands might moderate the relation between personality traits and conceptually related performance ratings both on the predictor and on the criterion side.

Future research is also needed to dig deeper into the relationship between situation assessment and personality. As noted in the Introduction, prior research provided inconsistent findings for this relationship, even when the same personality trait was considered. One approach might be to search for moderators of this relationship. One such moderator might be whether personality ratings are collected in an evaluative (e.g., applicant settings) versus research context. Specifically, if personality is measured in an evaluative context, the correlations between personality and situation assessment are higher (Klehe et al., in press) compared to when personality is measured in a non-evaluative context (Ingold, Kleinmann, König, & Melchers, 2011; Jansen, Lievens, & Kleinmann, 2011). This might be understood by the fact that in evaluative contexts personality scores might also reflect a social desirability/impression management component. As noted before, situation assessment might help to manage impressions in the right direction.

Finally, a strength of the current study is that we measured situation assessment with a test instead of a self-report measure. Accordingly, we did not measure how well people think they assess situational demands, but were able to measure how well they actually assessed the situational requirements. Thus, the present findings go beyond previous social effectiveness research that mainly focused on self-report measures (e.g., Ferris et al., 2005; Semadar, Robins, & Ferris, 2006; Zaccaro, Foti, & Kenny, 1991). Yet, more research is needed to investigate how self-report measures of social effectiveness and tests of candidates' assessment of situational demands like the one used in the present study are related to each other. In addition, the measurement of situation assessment also deserves further refinement in the future. In this study, we aggregated candidates' assessment of situational demands scores across performance criteria (dimensions) to increase the internal consistency reliability

of the situation assessment score. However, if situation assessment is measured on multiple occasions it might be possible to use reliable situation assessment scores per situational demand (performance criterion). This would make it feasible to test whether some performance criteria can be more easily deciphered than others and/or more easily translated into effective behavior (also see Table B1 in Appendix B).

Implications for Practice

In terms of practical implications, companies might consider using candidates' assessment of situational demands as part of a selection battery as our results showed that situation perception adds substantive variance to the prediction of job performance. Such a test could easily be administered by asking job applicants afterwards what they thought was assessed in the selection procedures (with a questionnaire or an interview) and would take only a few minutes. As we demonstrated that our measure is related to job performance, this test is job-related and might be used in making selection decisions provided that candidates are informed about it. In fact, assessing candidates' self-evaluations is already part of (developmental) ACs (Thornton & Rupp, 2006). Therefore, assessing candidates' perceptions of the situational demands of the selection process (i.e., the performance criteria targeted) might be of additional use for making selection decisions, particularly for jobs where it is essential to assess the situational requirements.

However, prior to implementing our suggestion to use candidates' assessment of situational demands as an additional "test" in the selection procedure, further research is needed. First, in the present study, participants were told that their assessment of situational demands score was used for research purposes only. We do not know what will happen to its validity when applicants are informed that their results are used for selection decisions. We expect that the scores of the test will be comparable when administered for selection purposes and research purposes because faking is a moot issue in ability measures. Yet, further

research should examine the consequences of using the test in a high-stakes context. Second, we need to examine potential subgroup differences of this test. There are two competing perspectives here. On one hand, subgroup differences might be anticipated in light of the correlation between cognitive ability and the assessment of situational demands, thereby potentially putting some subgroups at a disadvantage. On the other hand, it is also possible that limited subgroup differences exist in the assessment of situational demands measure. This might then explain why ACs and structured interviews display somewhat lower subgroup differences than cognitive ability (Hough, Oswald, & Ployhart, 2001; Ployhart & Holtz, 2008) because the former selection procedures also seem to indirectly measure the ability to assess situational demands.

Another issue that should be considered prior to implementing the assessment of situational demands as a test in selection practice concerns the coachability of this test. So far, we do not know whether and to what extent people can be coached in “assessing” situational demands and whether there exist subgroup differences in this coachability. Given that the assessment of situational demands is conceptualized as an ability, we expect that people might improve on it. As assessing situational demands is linked to job performance, it would be beneficial to know whether and how people can improve on their assessment.

Conclusion

In sum, this study presents new insights into the notion of behavioral consistency by demonstrating that individual differences in assessing situational demands in behavior-based selection procedures are significantly related to performance in these selection procedures and on the job. This study further discovered that people’s assessment of situational demands plays a key role in the criterion-related validity of selection procedures based on the behavioral consistency logic.

References

- Alexander, C. N., Jr., & Knight, G. W. (1971). Situated identities and social psychological experimentation. *Sociometry*, *34*, 65-82. doi:10.2307/2786351
- Alexander, C. N., Jr., & Rudd, J. (1984). Predicting behaviors from situated identities. *Social Psychology Quarterly*, *47*, 172-177. doi:10.2307/3033945
- Arthur, W., Jr., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology*, *56*, 125-154. doi:10.1111/j.1744-6570.2003.tb00146.x
- Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology*, *81*, 261-272. doi:10.1037/0021-9010.81.3.261
- Berry, C. M., Sackett, P. R., & Landers, R. N. (2007). Revisiting interview-cognitive ability relationships: Attending to specific range restriction mechanisms in meta-analysis. *Personnel Psychology*, *60*, 837-874. doi:10.1111/j.1744-6570.2007.00093.x
- Biderman, M. D., Nguyen, N. T., Mullins, B., & Luna, J. (2008, April). *A method factor predictor of performance ratings*. Paper presented at the 23rd annual conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.
- Binning, J. F., & Barrett, G. V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology*, *74*, 478-494. doi:10.1037/0021-9010.74.3.478
- Blickle, G., Momm, T., Schneider, P. B., Gansen, D., & Kramer, J. (2009). Does acquisitive self-presentation in personality self-ratings enhance validity? Evidence from two experimental field studies. *International Journal of Selection and Assessment*, *17*, 142-153. doi:10.1111/j.1468-2389.2009.00458.x

- Block, J., & Block, J. H. (1981). Studying situational dimensions: A grand perspective and some limited empiricism. In D. Magnussen (Ed.), *Toward a psychology of situations: An interactionist perspective* (pp. 85-102). Hillsdale, NJ: Erlbaum.
- Bott, J. P., Svyantek, D. J., Goodmann, S. A., & Bernal, D. S. (2003). Expanding the performance domain: Who says nice guys finish last? *International Journal of Organizational Analysis*, *11*, 137-152. doi:10.1108/eb028967
- Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2 ed., Vol. 1, pp. 687-732). Palo Alto, CA: Consulting Psychologists Press.
- Collins, J. M., Schmidt, F. L., Sanchez-Ku, M., Thomas, L., McDaniel, M. A., & Le, H. (2003). Can basic individual differences shed light on the construct meaning of assessment center evaluations? *International Journal of Selection and Assessment*, *11*, 17-29. doi:10.1111/1468-2389.00223
- Dierdorff, E. C., & Morgeson, F. P. (2007). Consensus in work role requirements: The influence of discrete occupational context on role expectations. *Journal of Applied Psychology*, *92*, 1228-1241. doi:10.1037/0021-9010.92.5.1228
- Dilchert, S., & Ones, D. S. (2009). Assessment center dimensions: Individual differences correlates and meta-analytic incremental validity. *International Journal of Selection and Assessment*, *17*, 254-270. doi:10.1111/j.1468-2389.2009.00468.x
- Ekehammar, B. (1974). Interactionism in personality from a historical perspective. *Psychological Bulletin*, *81*, 1026-1048. doi:10.1037/h0037457
- Ellingson, J. E., Smith, D. B., & Sackett, P. R. (2001). Investigating the influence of social desirability on personality factor structure. *Journal of Applied Psychology*, *86*, 122-133. doi:10.1037//0021-9010.86.1.122

- Endler, N. S., & Magnusson, D. (1976). Toward an interactional psychology of personality. *Psychological Bulletin*, *83*, 956-974. doi:10.1037/0033-2909.83.5.956
- Eurich, T. L., Krause, D. E., Cigularov, K., & Thornton, G. C., III. (2009). Assessment centers: Current practices in the United States. *Journal of Business and Psychology*, *24*, 387-407. doi:10.1007/s10869-009-9123-3
- Ferris, G. R., & Judge, T. A. (1991). Personnel/human resources management: A political influence perspective. *Journal of Management*, *17*, 447-488.
doi:10.1177/014920639101700208
- Ferris, G. R., Perrewé, P. L., & Douglas, C. (2002). Social effectiveness in organizations: Construct validity and research directions. *Journal of Leadership & Organizational Studies*, *9*, 49-63. doi:10.1177/107179190200900104
- Ferris, G. R., Treadway, D. C., Kolodinsky, R. W., Hochwarter, W. A., Kacmar, C. J., Douglas, C., & Frink, D. D. (2005). Development and validation of the political skill inventory. *Journal of Management*, *31*, 642-643. doi:10.1177/0149206304271386
- Fleeson, W. (2007). Situation-based contingencies underlying trait-content manifestation in behavior. *Journal of Personality*, *75*, 825-862. doi:10.1111/j.1467-6494.2007.00458.x
- Fox, S., & Spector, P. E. (2000). Relations of emotional intelligence, practical intelligence, general intelligence, and trait affectivity with interview outcomes: It's not all just "G". *Journal of Organizational Behavior*, *21*, 203-220. doi:10.1002/%28SICI%291099-1379%28200003%2921:2%3C203::AID-JOB38%3E3.0.CO;2-Z
- Funder, D. C. (2006). Towards a resolution of the personality triad: Persons, situations, and behaviors. *Journal of Research in Personality*, *40*, 21-34.
doi:10.1016/j.jrp.2005.08.003

- Hoefl, S., & Schuler, H. (2001). The conceptual basis of assessment centre ratings. *International Journal of Selection and Assessment, 9*, 114-123. doi:10.1111/1468-2389.00168
- Hogan, R., & Shelton, D. (1998). A socioanalytic perspective on job performance. *Human Performance, 11*, 129-144. doi:10.1207/s15327043hup1102&3_2
- 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. doi:10.1037/0021-9010.75.5.581
- Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence, and lessons learned. *International Journal of Selection and Assessment, 9*, 152-194. doi:10.1111/1468-2389.00171
- Huffcutt, A. I., Conway, J. M., Roth, P. L., & Stone, N. J. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology, 86*, 897-913. doi:10.1037/0021-9010.86.5.897
- Huffcutt, A. I., Roth, P. L., & McDaniel, M. A. (1996). A meta-analytic investigation of cognitive ability in employment interview evaluations: Moderating characteristics and implications for incremental validity. *Journal of Applied Psychology, 81*, 459-473. doi:10.1037/0021-9010.81.5.459
- Ingold, P., Kleinmann, M., König, C. J., & Melchers, K. G. (2011, June). *Scoring on the trait versus knowing what is being assessed: What explains the criterion-related validity of the interview?* Paper presented at the EAWOP small group meeting on selection and assessment, Athens, Greece.

International Task Force on Assessment Center Guidelines. (2009). Guidelines and ethical considerations for assessment center operations. *International Journal of Selection and Assessment, 17*, 243-253. doi:10.1111/j.1468-2389.2009.00467.x

Jansen, A., Lievens, F., & Kleinmann, M. (2011). Do individual differences in perceiving situational demands moderate the relationship between personality and assessment center dimension ratings? *Human Performance, 24*, 231-250.
doi:10.1080/08959285.2011.580805

Jansen, A., Melchers, K. G., & Kleinmann, M. (2011, September). *Beeinflusst soziale Kompetenz die Leistung im Assessment Center? Ein Vergleich verschiedener Maße sozialer Kompetenz [Does social skill influence assessment center performance? A comparison of different social skill measures]*. Paper presented at the 7th conference of the German Society for Work, Organizational and Business Psychology, Rostock, Germany.

Janz, T. (1989). The patterned behavior description interview: The best prophet of the future is the past. In R. W. Eder & G. R. Ferris (Eds.), *The employment interview: Theory, research, and practice* (pp. 158-168). Thousand Oaks, CA: Sage.

Klehe, U.-C., Kleinmann, M., Hartstein, T., Melchers, K. G., König, C. J., Heslin, P., & Lievens, F. (in press). Responding to personality tests in a selection context: The role of the ability to identify criteria and the ideal-employee factor. *Human Performance*.

Kleinmann, M. (1993). Are rating dimensions in assessment centers transparent for participants? Consequences for criterion and construct validity. *Journal of Applied Psychology, 78*, 988-993. doi:10.1037/0021-9010.78.6.988

Kleinmann, M. (1997). Transparenz der Anforderungsdimensionen: Ein Moderator der Konstrukt- und Kriteriumsvalidität des Assessment-Centers [Transparency of the

- required dimensions: A moderator of assessment centers' construct and criterion validity]. *Zeitschrift für Arbeits- und Organisationspsychologie*, *41*, 171-181.
- Kleinmann, M., Ingold, P. V., Lievens, F., Jansen, A., Melchers, K. G., & König, C. J. (2011). A different look at why selection procedures work. *Organizational Psychology Review*, *1*, 128-146. doi:10.1177/2041386610387000
- Kleinmann, M., & Klehe, U.-C. (2011). Selling oneself: Construct and criterion-related validity of impression management in structured interviews. *Human Performance*, *24*, 29-46. doi:10.1080/08959285.2010.530634
- Kolk, N. J., Born, M. P., & van der Flier, H. (2003). The transparent assessment centre: The effects of revealing dimensions to candidates. *Applied Psychology: An International Review*, *52*, 648-668. doi:10.1111/1464-0597.00156
- König, C. J., Melchers, K. G., Kleinmann, M., Richter, G. M., & Klehe, U.-C. (2007). Candidates' ability to identify criteria in nontransparent selection procedures: Evidence from an assessment center and a structured interview. *International Journal of Selection and Assessment*, *15*, 283-292. doi:10.1111/j.1744-6570.1987.tb00603.x
- Latham, G. P. (1989). The reliability, validity, and practicality of the situational interview. In R. W. Eder & G. R. Ferris (Eds.), *The employment interview: Theory, research, and practice* (pp. 169-182). Thousand Oaks, CA: Sage.
- Latham, G. P., & Sue-Chan, C. (1999). A meta-analysis of the situational interview: An enumerative review of reasons for its validity. *Canadian Psychology*, *40*, 56-67. doi:10.1037/h0086826
- Levashina, J., & Campion, M. A. (2006). A model of faking likelihood in the employment interview. *International Journal of Selection and Assessment*, *14*, 299-316. doi:10.1111/j.1468-2389.2006.00353.x

- Lievens, F., Sanchez, J. I., Bartram, D., & Brown, A. (2010). Lack of consensus among competency ratings of the same occupation: Noise or substance? *Journal of Applied Psychology, 95*, 562-571. doi:10.1037/a0018035
- Maurer, S. D., Sue-Chan, C., & Latham, G. P. (1999). The situational interview. In R. W. Eder & M. M. Harris (Eds.), *The employment interview handbook* (pp. 159-177). Thousand Oaks, CA: Sage.
- McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. L., III. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology, 60*, 63-91. doi:10.1111/j.1744-6570.2007.00065.x
- McFarland, L. A., Yun, G., Harold, C. M., Viera, L., Jr., & 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. doi:10.1111/j.1744-6570.2005.00374.x
- Melchers, K. G., Bösser, D., Hartstein, T., & Kleinmann, M. (in press). Assessment of situational demands in a selection interview: Reflective style or sensitivity? . *International Journal of Selection and Assessment*.
- Melchers, K. G., Kleinmann, M., Richter, G. M., König, C. J., & Klehe, U.-C. (2004). Messen Einstellungsinterviews das, was sie messen sollen? Zur Bedeutung der Bewerberkognitionen über bewertetes Verhalten [Do selection interviews measure what they intend to measure? The impact of applicants' cognitions about evaluated behavior]. *Zeitschrift für Personalpsychologie, 3*, 159-169. doi:doi:10.1026/1617-6391.3.4.159
- 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. doi:10.1037/0021-9010.93.5.1042
- 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. doi:10.1177/0149206309349309
- 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. doi:10.1037/0033-295X.102.2.246
- Mischel, W., & Shoda, Y. (1998). Reconciling processing dynamics and personality dispositions. *Annual Review of Psychology*, *49*, 229-258. doi:10.1146/annurev.psych.49.1.229
- Morrison, E. W. (1993). Newcomer information seeking: Exploring types, modes, sources, and outcomes. *Academy of Management Journal*, *36*, 557-589. doi:10.2307/256592
- Motowidlo, S. J. (1999). Asking about past behavior versus hypothetical behavior. In R. W. Eder & M. M. Harris (Eds.), *The employment interview handbook* (pp. 179-190). Thousand Oaks, CA: Sage.
- Motowidlo, S. J., Borman, W. C., & Schmit, M. J. (1997). A theory of individual differences in task and contextual performance. *Human Performance*, *10*, 71-83. doi:10.1207/s15327043hup1002_1
- Murphy, K. R. (1989). Is the relationship between cognitive ability and job performance stable over time? *Human Performance*, *2*, 183-200. doi:10.1207/s15327043hup0203_3
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, *81*, 660-679. doi:10.1037/0021-9010.81.6.660

- Pervin, L. A. (1989). Persons, situations, interactions: The history of a controversy and a discussion of theoretical models. *Academy of Management Review*, *14*, 350-360. doi:10.2307/258172
- Ployhart, R. E., & Holtz, B. C. (2008). The diversity-validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, *61*, 153-172. doi:10.1111/j.1744-6570.2008.00109.x
- Preckel, D., & Schüpbach, H. (2005). Zusammenhänge zwischen rezeptiver Selbstdarstellungskompetenz und Leistung im Assessment Center [Correlations between receptive self-presentation competence and performance in an assessment center]. *Zeitschrift für Personalpsychologie*, *4*, 151-158. doi:10.1026/1617-6391.4.4.151
- Reis, H. T. (2008). Reinvigorating the concept of situation in social psychology. *Personality and Social Psychology Review*, *12*, 311-329. doi:10.1177/1088868308321721
- Roth, P. L., Bobko, P., & McFarland, L. A. (2005). A meta-analysis of work sample test validity: Updating and integrating some classic literature. *Personnel Psychology*, *58*, 1009-1037. doi:10.1111/j.1744-6570.2005.00714.x
- Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., & de Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities: A European community meta-analysis. *Personnel Psychology*, *56*, 573-605. doi:10.1111/j.1744-6570.2003.tb00751.x
- Salgado, J. F., & Moscoso, S. (2002). Comprehensive meta-analysis of the construct validity of the construct-validity of the employment interview. *European Journal of Work and Organizational Psychology*, *11*, 299-324. doi:10.1080/13594320244000184

- Schleicher, D. J., & Day, D. V. (1998). A cognitive evaluation of frame-of-reference rater training: Content and process issues. *Organizational Behavior and Human Decision Processes*, *73*, 76-101. doi:10.1006/obhd.1998.2751
- Schmitt, N., & Ostroff, C. (1986). Operationalizing the "behavioral consistency" approach: Selection test development based on a content-oriented strategy. *Personnel Psychology*, *39*, 91-108. doi:10.1111/j.1744-6570.1986.tb00576.x
- Schmitt, N., & Oswald, F. L. (2006). The impact of corrections for faking on the validity of noncognitive measures in selection settings. *Journal of Applied Psychology*, *91*, 613-621. doi:10.1037/0021-9010.91.3.613
- Schulz-Hardt, S., Mojzisch, A., & Brodbeck, F.-C. (2006). Group decision making in hidden profile situations: Dissent as a facilitator for decision quality. *Journal of Personality and Social Psychology*, *91*, 1080-1093. doi:10.1037/0022-3514.91.6.1080
- Semadar, A., Robins, G., & Ferris, G. R. (2006). Comparing the validity of multiple social effectiveness constructs in the prediction of managerial job performance. *Journal of Organizational Behavior*, *27*, 443-461. doi:10.1002/job.385
- Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, *30*, 526-537.
- Speer, A. B., Christiansen, N. D., Melchers, K. G., König, C. J., & Kleinmann, M. (2012, April). *Cross-situational convergence of ability to identify criteria across AC exercises*. Paper presented at the 27th annual conference of the Society of Industrial and Organizational Psychology, San Diego, CA.
- Spychalski, A. C., Quiñones, M. A., Gaugler, B. B., & Pohley, K. (1997). A survey of assessment center practices in organizations in the United States. *Personnel Psychology*, *50*, 71-90. doi:10.1111/j.1744-6570.1997.tb00901.x

- Staufenbiel, T., & Hartz, C. (2000). Organizational citizenship behavior: Entwicklung und erste Validierung eines Messinstruments. [Organizational citizenship behavior: Development and validation of a measurement instrument]. *Diagnostica, 46*, 73-83. doi:10.1026//0012-1924.46.2.73
- Ten Berge, M. A., & De Raad, B. (1999). Taxonomies of situations from a trait psychological perspective. A review. *European Journal of Personality, 13*, 337-360. doi:10.1002/(sici)1099-0984(199909/10)13:5<337::aid-per363>3.0.co;2-f
- Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology, 88*, 500-517. doi:10.1037/0021-9010.88.3.500
- Thorndike, E. L. (1920). Intelligence and its use. *Harper's Magazine, 140*, 227-235.
- Thornton, G. C., III, & Rupp, D. E. (2006). *Assessment centers in human resource management: Strategies for prediction, diagnosis, and development*. Mahwah, NJ: Erlbaum.
- Tullar, W. L. (1989). The employment interview as a cognitive performing script. In R. W. Eder & G. R. Ferris (Eds.), *The employment interview: Theory, research, and practice* (pp. 233-245). Newbury Park, CA: Sage.
- Viswesvaran, C., & Ones, D. S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment, 8*, 216-226. doi:10.1111/1468-2389.00151
- Wernimont, P. F., & Campbell, J. T. (1968). Signs, samples, and criteria. *Journal of Applied Psychology, 52*, 372-376. doi:10.1037/h0026244
- Williams, L. J., & Anderson, S. E. (1991). Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. *Journal of Management, 17*, 601-617. doi:10.1177/014920639101700305

Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, *67*, 189-205.

Wonderlic Inc. (2002). *Wonderlic Personnel Test and Scholastic Level Exam: User's manual*. Libertyville, IL.: Author.

Zaccaro, S. J., Foti, R. J., & Kenny, D. A. (1991). Self-monitoring and trait-based variance in leadership: An investigation of leader flexibility across multiple group situations. *Journal of Applied Psychology*, *76*, 308-315. doi:10.1037/0021-9010.76.2.308

Table 1

Correlates of Assessment of Situational Demands in Previous Studies

Source	<i>r</i>	<i>N</i>	Correlate
Cognitive Ability			
Klehe et al. (in press)	.17	149	General mental ability
König et al. (2007, situation assessment in the interview)	.30	94	Verbal ability
	.12	94	Figural ability
König et al. (2007, situation assessment in the AC)	.29	94	Verbal ability
	.16	94	Figural ability
Kolk, Born, and van der Flier (2003)	.27	149	Verbal ability
Preckel and Schüpbach (2005)	.29	71	Reasoning
Personality			
Ingold Kleinmann, König, and Melchers (2011)	.02	99	Persistence
	.13	99	Assertiveness
	-.14	99	Organizing behavior
Jansen, Lievens and Kleinmann (2011)	.04	108	Openness to experience
	-.03	108	Conscientiousness
	.02	108	Extraversion
	.08	108	Agreeableness
	.08	108	Emotional stability
Klehe et al. (in press)	.17	149	Openness to experience
	.30	149	Conscientiousness
	.12	149	Extraversion
	.26	149	Agreeableness
	.24	149	Emotional stability
Preckel and Schüpbach (2005)	-.19	71	Conscientiousness
	.06	71	Extraversion
	-.23	71	Agreeableness
	.33	71	Emotional stability
Social Effectiveness			
Klehe et al. (in press)	.08	149	Self-monitoring
Kleinmann (1997)	.26	62	Social perceptiveness
Jansen, Melchers, and Kleinmann (2011)	.25	114	Political skill
	.01	114	Self-monitoring
Interview Performance			
Ingold et al. (2011)	.23	99	Interview performance
Klehe et al. (in press)	.26	149	Interview performance
König et al. (2007, situation assessment in the interview)	.35	94	Interview performance
König et al. (2007, situation assessment in the AC)	.29	94	Interview performance
Melchers, Kleinmann, Richter, König, and Klehe (2004)	.27	62	Interview performance
AC Performance			
Speer, Christiansen, Melchers, König, and Kleinmann (2012)	.40	173	AC performance
Jansen, Lievens, and Kleinmann (2011)	.23	108	AC performance
Jansen, Melchers, and Kleinmann (2011)	.25	114	AC performance

Klehe et al. (in press)	.27	149	AC performance
Kleinmann (1993)	.30	56	AC performance
Kleinmann (1997)	.32	62	AC performance
König et al. (2007, situation assessment in the interview)	.34	94	AC performance
König et al. (2007, situation assessment in the AC)	.39	94	AC performance
Preckel and Schüpbach (2005)	.49	71	AC performance

Table 2
Descriptive Statistics and Intercorrelations of Study Variables

Variables	<i>M</i>	<i>SD</i>	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.
1. Sex	1.46	0.50																						
2. Age	27.61	4.14	.03																					
3. Cognitive ability	31.71	5.72	-.19*	-.03																				
4. Situation assessment	1.54	0.40	.09	-.18*	.29*																			
5. Organizing (SA)	1.50	0.75	.06	-.08	.11	.64*																		
6. Presentation skills (SA)	2.21	0.60	.09	-.05	.24*	.23*	-.03																	
7. Creativity (SA)	1.02	1.29	-.09	-.08	.04	.43*	.03	-.05																
8. Analytical skills (SA)	0.80	0.73	.02	-.09	.07	.37*	.07	-.06	.20*															
9. Persuasiveness (SA)	1.59	0.80	-.01	-.14	.25*	.56*	.12	.05	.06	.08														
10. Assertiveness (SA)	1.44	1.26	.01	.08	.06	.41*	.14	-.03	.20*	.12	.22*													
11. Consideration (SA)	1.82	0.87	.19*	-.12	.14	.42*	.11	.20*	.18*	-.02	.06	-.06												
12. Organizing (AC)	3.22	0.64	-.21*	.07	.26*	.21*	.29*	.12	.14	.03	-.06	-.07	.14											
13. Presentation skills (AC)	3.53	0.78	-.23*	-.08	.26*	.15	.17	.09	.06	-.04	.02	.03	.12	.63*										
14. Creativity (AC)	3.02	1.21	-.15	.03	-.01	.16	.17	.04	.20*	-.03	-.03	.14	.06	.25*	.26*									
15. Analytical skills (AC)	3.39	0.79	-.22*	-.02	.33*	.22*	.14	.17	.09	.03	.12	.01	.16	.57*	.46*	.06								
16. Persuasiveness (AC)	3.33	0.72	-.26*	.01	.34*	.24*	.24*	.19*	.11	.06	.01	.06	.11	.65*	.63*	.18*	.75*							
17. Assertiveness (AC)	3.02	1.04	-.07	.05	.07	.10	.11	.01	.02	-.07	-.02	.18	.11	.38*	.23*	-.09	.41*	.49*						
18. Consideration (AC)	3.31	0.56	-.13	-.02	.08	.17	.15	.14	.03	.03	.06	-.10	.18*	.54*	.47*	.17	.50*	.54*	.26*					
19. Organizing (Interview)	3.40	0.68	-.09	-.03	.22*	.24*	.20*	-.03	.11	.06	.19*	.02	.09	.43*	.20*	.08	.18*	.21*	.01	.28*				
20. Consideration (Interview)	3.60	0.63	.12	.05	.16	.21*	.14	.13	-.01	.03	.18*	-.08	.15	.29*	.21*	.06	.13	.19*	.03	.47*	.58*			
21. Overall interview rating	3.50	0.58	-.01	.01	.22*	.26*	.19*	.06	.06	.05	.21*	-.03	.13	.41*	.23*	.08	.18	.23*	.03	.42*	.90*	.88*		
22. Overall AC rating	3.26	0.58	-.28*	.02	.27*	.23*	.26*	.16	.14	.00	-.02	.02	.17	.86*	.77*	.38*	.75*	.86*	.50*	.68*	.22*	.20	.24*	
23. Job performance ^a	5.90	0.85	-.11	-.14	.18	.27*	.10	-.04	.10	.14	.18	.19	.24*	.19	.25*	.19*	.17	.15	.05	.13	.19	.15	.19*	.21*

Note. $N = 124$ with the exception of job performance where $N = 107$. Sex was coded 1 = male, 2 = female. SA = situation assessment, AC = assessment center. ^a because of rounding to two decimal places, some of the correlations with job performance for which a value of .19 is given are significant whereas other are not.

* $p < .05$ (two-tailed).

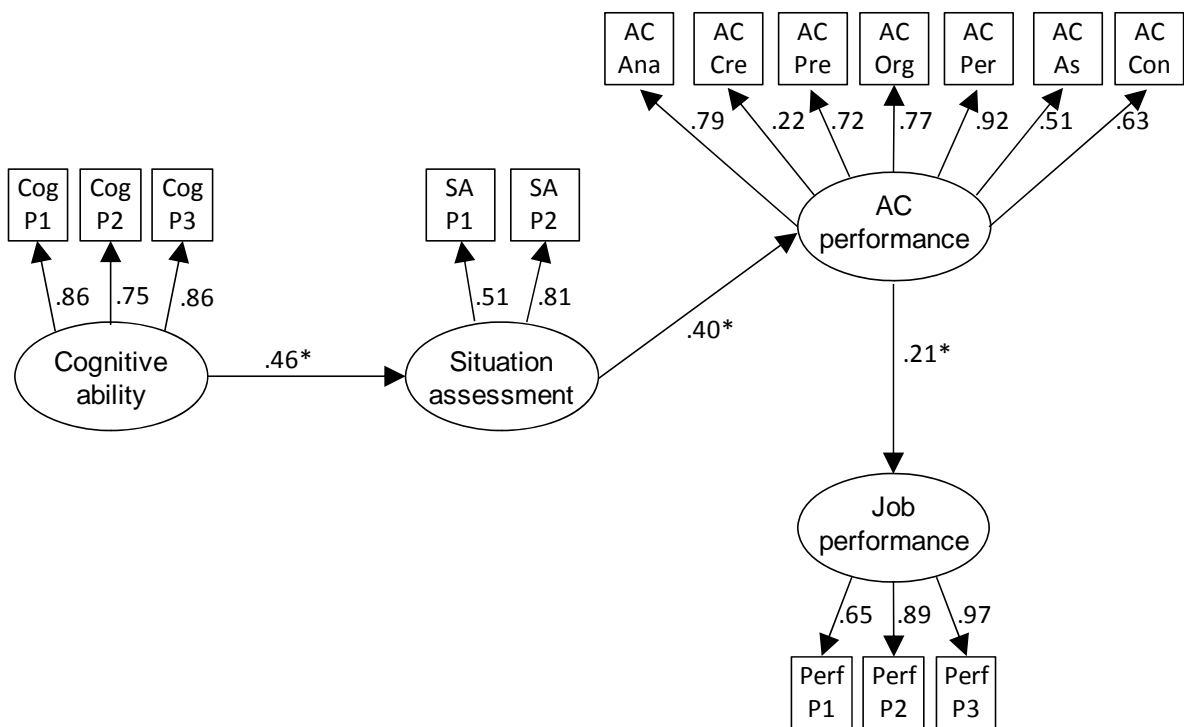
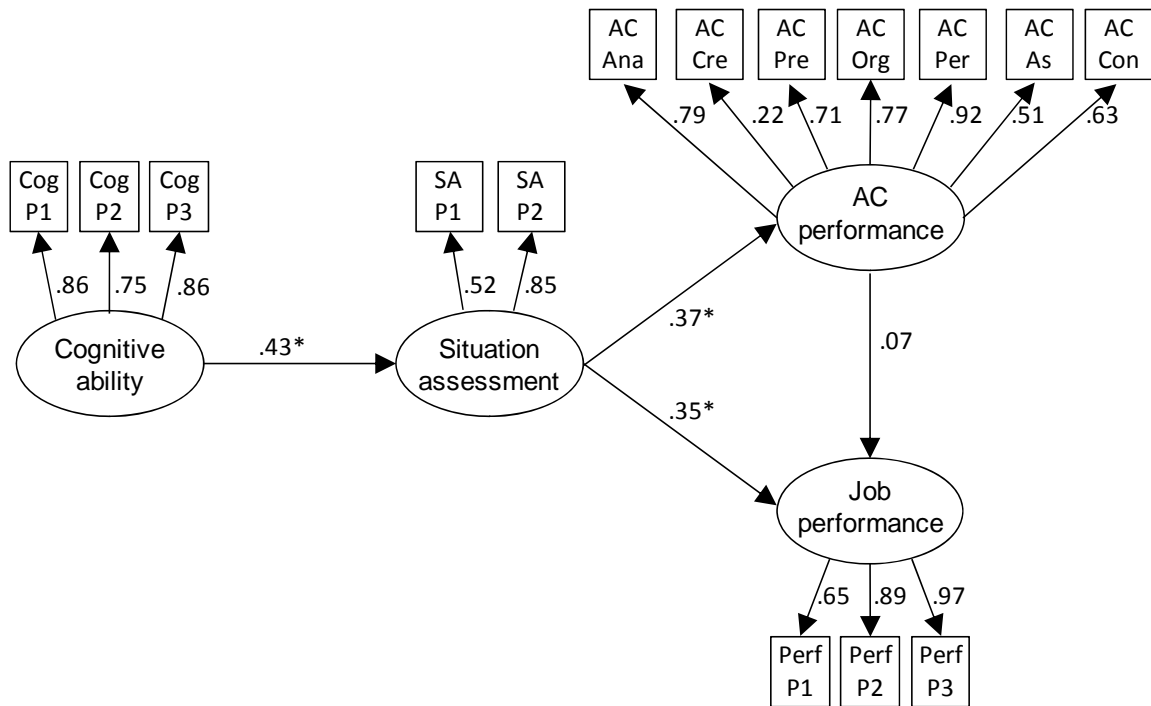


Figure 1. Standardized parameter estimates of the hypothesized model. To simplify the figure, all error terms were omitted. Cog = cognitive ability, SA = situation assessment, AC = assessment center, Perf = job performance, P = item parcel, Ana = analytical skills, Cre = creativity, Pre = presentation skills, Org = organizing, Per = persuasiveness, As = assertiveness, and Con = consideration of others. * $p < .05$.

Appendix A

Description of AC Exercises and Interview

The AC exercises comprised two leaderless group discussions (four to six participants) and two presentations. The first group discussion exercise represented a hidden-profile task (i.e., a task with an asymmetrical distribution of information, Schulz-Hardt, Mojzisch, & Brodbeck, 2006). Participants had to discuss a staffing task and had to collaborate with each other to identify the most qualified candidate for a vacant position. To find the best solution, they needed to discover that they held different pieces of information and had to share this information. In the second group discussion exercise, participants were first asked to rank-order eight counterproductive behaviors on their own and then to find a collective solution in the group. They were instructed that the collective rank order should be as similar as possible to their individual rank order. In the first presentation exercise, the participants received information materials about the manufacturing system of a fictitious company and had to prepare and conduct a presentation to convince a potential customer of the value of the product. In the other presentation exercise, they received five minutes to introduce themselves to the assessors. The interview consisted of two different components: six past-oriented questions (cf. Janz, 1989; Motowidlo, 1999) and six future-oriented questions (cf. Latham, 1989).

All exercises and the interview questions were pre-tested with groups of I/O psychology graduate students. The pre-tests assessed whether exercise instructions and interview questions were understandable and how well exercises measured the dimensions targeted in contrast to various other dimensions.

Appendix B

Table B1

Situational Assessment Scores from Previous Studies for Different Categories of Dimensions. To Sort Dimensions into such Categories, the Category System from Arthur et al. (2003) is Used. All Scores are Expressed as a Value between 0-1; for Studies with More than 1 Dimension in the Respective Category Only the Mean Value is Given.

Study, selection procedure used and <i>N</i>	Communi- cation	Consideration/ Awareness of Others	Drive	Influencing Others	Organizing & Planning	Problem Solving	Tolerance for Stress/ Uncertainty
Ingold et al. (2011) interview (<i>N</i> = 109)		.46		.28	.19		
Jansen, Lievens, and Kleinmann (2011) AC (<i>N</i> = 108)	.68	.32		.43	.40	.25	.48
Jansen, Melchers, and Kleinmann (2011) AC (<i>N</i> = 117)	.67	.53		.53	.38	.28	
Klehe et al. (in press) AC (<i>N</i> = 147)		.38		.42	.55		
Klehe et al. (in press) interview (<i>N</i> = 147)		.55		.53	.66		
Kleinmann (1993) AC (<i>N</i> = 56)		.45		.53	.60		
König et al. (2007) AC (<i>N</i> = 95)		.37		.37	.50		
König et al. (2007) interview (<i>N</i> = 95)		.44		.47	.56		
Melchers et al. (2004) interview (<i>N</i> = 64)	.48		.52	.57	.46		
Sample-weighted mean	.62	.43	.52	.45	.47	.27	.48
Overall <i>N</i>	237	829	109	938	938	173	56
<i>k</i>	3	8	1	9	9	2	1

Note. *k* = number of studies on which a sample-weighted mean is based.