

# Journal of Applied Psychology

## **A Dual-Process Theory Perspective to Better Understand Judgments in Assessment Centers: The Role of Initial Impressions for Dimension Ratings and Validity**

Pia V. Ingold, Mirjam Dönni, and Filip Lievens

Online First Publication, July 30, 2018. <http://dx.doi.org/10.1037/apl0000333>

### CITATION

Ingold, P. V., Dönni, M., & Lievens, F. (2018, July 30). A Dual-Process Theory Perspective to Better Understand Judgments in Assessment Centers: The Role of Initial Impressions for Dimension Ratings and Validity. *Journal of Applied Psychology*. Advance online publication. <http://dx.doi.org/10.1037/apl0000333>

## RESEARCH REPORT

A Dual-Process Theory Perspective to Better Understand Judgments in  
Assessment Centers: The Role of Initial Impressions for Dimension  
Ratings and ValidityPia V. Ingold  
University of ZurichMirjam Dönni  
Aviva Selection, Basel, SwitzerlandFilip Lievens  
Singapore Management University

Insight into assessors' initial impressions has the potential to advance knowledge on how assessors form dimension-based judgments and on possible biases in these ratings. Therefore, this study draws on dual process theory to build and test a model that integrates assessors' dimension ratings (i.e., systematic, slow, deliberate processing mode) with their initial impressions (i.e., intuitive, fast, automatic processing mode). Data collection started with an AC where assessors provided ratings of assesseees, and an online survey of assesseees' supervisors who rated their job performance. In addition, two other rater pools provided initial impressions of these assesseees by evaluating extracted 2-min video clips of their AC performance. Initial impressions from both of these samples were positively related to assessors' dimension ratings, which supports assumptions from dual process theory and might explain why assessors' dimensional ratings are often undifferentiated. Initial impressions did not appear to open up the doors for biases and stereotypes based upon appearance and perceptions of liking. Instead, assessors picked up information that assesseees transmitted about their personality (i.e., Conscientiousness and Emotional Stability). Implications for further research on initial impressions and AC dimension ratings are discussed.

*Keywords:* assessment center, initial impression, dual process theory, rating process, criterion-related validity

In interpersonal personnel selection procedures such as employment interviews and assessment centers, decisions about candidates are based upon on human judgments, and these judgments play a critical role in selection decisions. For assessment centers (ACs), judgments are especially relevant given that a core char-

acteristic of ACs is that assessors evaluate candidates' behavior on dimensions (or other constructs of interest) in several simulated fast-paced interpersonal situations. As these evaluations lie at the heart of ACs, a key question concerns the kind of judgment processes underlying assessors' ratings. AC guidelines and textbooks (e.g., Rupp et al., 2015; Thornton & Rupp, 2006) typically suggest that systematic processes are operating when assessors observe behaviors, note down observations, and classify these observations into dimensions before evaluating the candidates.

However, dual process theories, that have made strong inroads to explain judgment processes in social and cognitive psychology (see Evans, 2008 for an overview), offer a promising, more comprehensive perspective on AC judgment processes. Dual process theories highlight that a systematic deliberate route (as traditionally proposed in the AC literature) constitutes only one of two *complementary* cognitive processing modes. According to dual process theories, we process information systematically, deliberately, slowly, and elaborately (also known as controlled mode of cognition, Type II, or System 2, Evans & Stanovich, 2013), but we also quickly arrive at conclusions via fast and frugal judgments (also known as the intuitive mode of cognition, Type I, or System

---

Pia V. Ingold, Department of Psychology, University of Zurich; Mirjam Dönni, Aviva Selection, Basel, Switzerland; Filip Lievens, Lee Kong Chian School of Business, Singapore Management University.

Parts of the research reported in this grant were supported by the Swiss National Science Foundation (grant 100014-124449). Parts of this research were presented at the 31st Annual Conference of the Society of Industrial and Organizational Psychology in Anaheim, California, 2016. We are indebted to Brian S. Connelly and Cornelius J. König for their helpful feedback on a previous version of this article. We thank Nadine Maier and Marion Hauert for helping collect the data.

Correspondence concerning this article should be addressed to Pia V. Ingold, Arbeits- und Organisationspsychologie, Universität Zürich, Binzmühlestrasse 14/12, CH-8050 Zürich, Switzerland. E-mail: p.ingold@psychologie.uzh.ch

1, Evans & Stanovich, 2013). Notably, in cognitively demanding situations like ACs, fast and frugal judgments are supposed to be especially influential for evaluation outcomes. In line with this, it has been recommended to consider the automatic mode for better understanding judgments in ACs (Highhouse, 1997; Zedeck, 1986). Yet, the extensive research base on dimension ratings attests that only the controlled judgment mode has received attention.

Drawing upon dual process theories, this article's premise is that insights into assessors' initial impressions (i.e., snap judgments made within the first minutes; Barrick, Swider, & Stewart, 2010) produce a more complete understanding of how assessors arrive at dimension ratings and contributes to at least two bodies of AC knowledge. First, studying assessors' initial impressions informs the research base on how assessors form dimension ratings. This is relevant because a robust finding is that assessors have difficulty providing differentiated judgments on AC dimensions (e.g., Bowler & Woehr, 2006; Jackson, Michaelides, Dewberry, & Kim, 2016). Yet, the actual AC rating process has remained largely uncovered territory, and unfortunately, Zedeck's (1986) 30-year old conclusion of "an almost total lack of research on process issues in assessment center methods" (p. 293) still stands. One exception is a study by Lance, Foster, Gentry, and Thoresen (2004) who found that a general impression model best explained the high intercorrelations among dimension ratings. Yet, Lance and colleagues did not examine what drove this general impression. On the basis of dual process theories, an untested hypothesis is that assessors' initial impressions serve as a first, general anchor that affects their dimension ratings. Second, a closer look at assessors' initial impressions contributes to research on potential biases in dimension ratings. Although AC ratings are less prone to subgroup differences than other selection procedures (e.g., cognitive ability tests), ethnic and sex differences are not negligible (Bobko & Roth, 2013; Dean, Roth, & Bobko, 2008). Based on dual process theories, an unexplored hypothesis is that initial impressions, quickly made on the basis of limited and salient information, carry biases that affect subsequent dimension ratings.

To answer these questions and advance AC knowledge, we build and test a model that integrates assessors' systematic dimension-based

judgments with their initial impressions. As such, this model offers a much-needed integration between the two modes of judgment that have been too often examined separately in research (Evans & Stanovich, 2013; Kruglanski & Gigerenzer, 2011). Through testing this model, we provide insights on (a) whether and how initial impressions are related to dimension ratings (e.g., analytical skills, cooperation, persuasiveness, presentation skills, organizing and planning), and (b) whether initial impressions introduce potentially biasing (i.e., perceptions of liking and attractiveness) or valid (i.e., perceptions of expressed personality) information into dimension ratings that may hinder or help the prediction of performance.

## Theoretical Background and Hypotheses

### Is There a Link Between Initial Impressions and AC Dimension Ratings?

The core part of our proposed model (see Figure 1) links initial impressions to AC dimension ratings. As noted above, Lance et al. (2004) conducted one of the sole investigations into judgment models underlying dimension ratings. To explain assessors' undifferentiated dimension judgments, they proposed two models: a salient dimension model and a general impression model. According to the salient dimension model, assessors view one or more dimensions as crucial, and these salient dimensions influence assessors' ratings on other dimensions. The general impression model posits that assessors form a general impression of assesseees, which then influences how they rate the AC dimensions (Lance, Foster, et al., 2004). Most support was found for the general impression model. However, Lance et al. collected assessors' general impressions after the dimension ratings, and left open the key question about the role of initial impressions for dimension ratings. This is where dual process theory helps to shed light on AC judgment processes and provides the fundament of our proposed model. One axiom of dual process theory posits that one spontaneously forms impressions on the basis of information that is readily available in the first couple of minutes. Another axiom posits

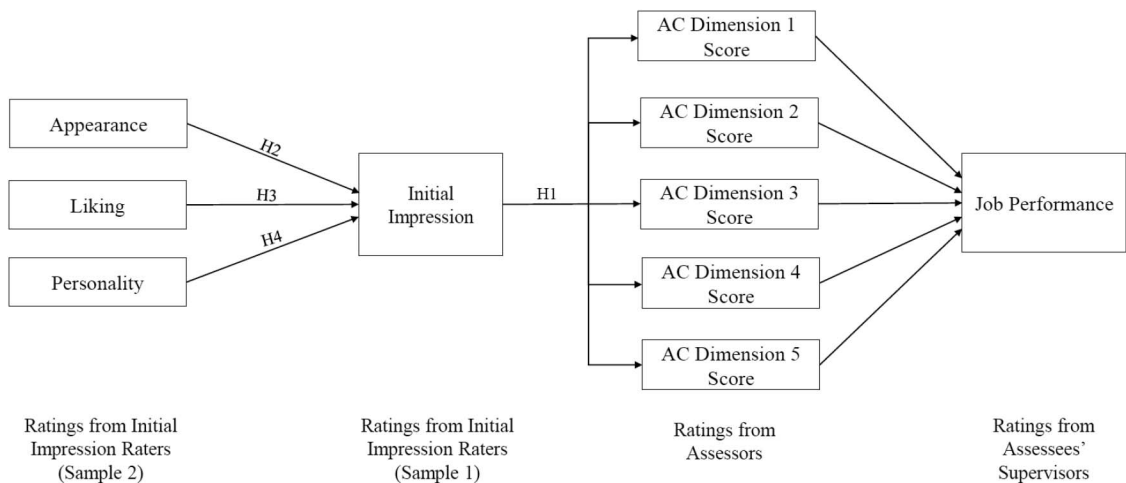


Figure 1. Overview of the research model on the role of initial impressions for AC dimension ratings and validity. Hypotheses (Hypothesis 1 (H1) to Hypothesis 4 (H4)) are displayed above arrows. Rating sources from the four samples are reported below the research model.

that such early information can be either endorsed, adjusted, or rejected by information later available. According to Kahneman (2003, p. 716) “highly accessible impressions produced by System 1 control judgments and preferences, unless modified or overridden by the deliberate operations of System 2” (see also Evans, 2008; Evans & Stanovich, 2013). Moreover, endorsing or adjusting early information seems more prevalent than rejecting it, especially under the conditions of cognitive load that assessors face (Kahneman, 2003). Thus, we expect that assessors’ initial impressions from the beginning of an exercise affect the systematic dimension ratings made at the end of the AC exercise. Inherent in the reasoning above is that initial impressions might also explain why assessors typically fail to distinguish between the various dimensions (e.g., Jackson et al., 2016) because these impressions serve as a first, global anchor for the later AC dimension ratings. Thus,

*Hypothesis (H1):* Initial impressions of assesseees are significantly related to dimension ratings.

Although the above dual process perspective suggests that assessors’ initial impressions serve as a first, general anchor for making dimension ratings at the end of the exercise, it is still important to explore whether all dimensions are equally affected. In fact, AC construct-related validity research suggests that the degree to which dimensions are affected might differ across dimensions (Bowler & Woehr, 2006), which would imply that some dimension ratings might be more affected by initial impressions than others. To investigate this more comprehensively, we drew from an AC dimension taxonomy (Meriac, Hoffman, & Woehr, 2014) that distinguishes between two broad types of AC dimensions: relational (e.g., consideration) and administrative dimensions (e.g., organization and planning). Hence, this framework is useful for exploring whether initial impressions are differentially related to these two types of AC dimensions. For instance, due to the interpersonal nature of initial impressions (see below), these impressions might affect ratings on relational dimensions more than ratings on administrative dimensions. Thus,

*Research Question (RQ1):* Are there differences in the relationship between initial impressions and administrative dimensions (analytical skills, organizing and planning) and the relationship between initial impressions and relational dimensions (persuasiveness, presentation skills, cooperation)?

### What Kind of Perceptions Inform Initial Impressions?

The second part of our model includes hypotheses of what kind of perceptions go into assessors’ initial impressions. Dual process theory assumes that automatic, intuitive judgment relies especially on quickly available information. The idea of “What you see is all there is” suggests that initial impressions are formed on the basis of readily available information, even if further key information is still lacking at that point (Fiske & Taylor, 1991; Kahneman, 2011). This part of the model posits that the following three pieces of information might be especially available to assessors early on. First, as visual attributes are readily available (Rule & Ambady, 2008b), we expect appearance (i.e., professional appearance and physical attractiveness) to play a prominent role in initial impressions. Besides dual process theories, various other theories (see Langlois et al., 2000 for an overview) also posit that one quickly develops positive expectancies of attractive and

professionally dressed people. Interview research further confirms that appearance is related to initial impressions (Swider, Barrick, & Harris, 2016). Thus,

*H2:* Perceptions of appearance are significantly related to initial impressions.

Second, we expect affective reactions such as perceptions of liking to be quickly available. Liking has been typified as a fundamental interpersonal perception dimension (Hartley et al., 2016; Wojciszke, Abele, & Baryla, 2009). This is because perceptions of liking are instantly formed in interpersonal encounters (Ambady & Rosenthal, 1992) and impact evaluation (Barrick et al., 2010; Sutton, Baldwin, Wood, & Hoffman, 2013). Hence, perceived liking is likely to be part of the information that goes into assessors’ initial impressions.

*H3:* Perceptions of liking are significantly related to initial impressions.

Third, interactional AC exercises provide assesseees with opportunities to express personality-related behaviors (Christiansen, Hoffman, Lievens, & Speer, 2013; Heimann, Ingold, Kleinmann, Lievens, & Melchers, 2017) and we expect these early available personality perceptions to inform assessors’ initial impressions. Our expectation of personality saturation in initial impressions is based on the large “zero-acquaintance” literature (Funder, 1999; Human & Biesanz, 2013; Kenny, 2004). In this zero-acquaintance paradigm, untrained people are asked to judge strangers on the basis of short excerpts of behavior (from less than 1 min to up to 5 min). Raters seem to quickly form impressions of people’s personality in these “thin-slices” of information situations, followed later by more holistic impressions (e.g., competence; Bar, Neta, & Linz, 2006; Willis & Todorov, 2006). This suggests that assessors’ initial impressions capture perceptions of assesseees’ expressed personality and that these perceptions also affect their later ratings.

To determine which Big Five trait perceptions would relate most strongly to assessors’ initial impressions, we formulated hypotheses on the basis of expectations about traits that people are most likely to self-promote in selection. To this end, we drew on research regarding applicant versus nonapplicant differences on the Big Five traits in personality inventories. According to a meta-analysis by Birkeland, Manson, Kisamore, Brannick, and Smith (2006), effect sizes are highest for Conscientiousness and Emotional Stability ( $d_s = .45$  and  $.46$ , respectively vs.  $d_s$  from  $.11$  to  $.16$  for the other traits). So, we expect that assesseees will aim to especially self-promote these two traits and thus express behaviors related to these traits in the first minutes of the AC exercises<sup>1</sup>, thereby influencing assessors’ initial impressions. Hence,

*H4:* Perceptions of expressed Conscientiousness (H4a) and Emotional Stability (H4b) will be significantly related to initial impressions.

<sup>1</sup> As noted by an anonymous reviewer, distorting responses on trait-related statements in self-report inventories might not always translate into behavioral expressions of these traits in AC exercises.

## Do Initial Impressions Affect Dimension Ratings? Criterion-Related Validity?

The last part of the model deals with the effects of initial impressions on dimension ratings and their criterion-related validity. A pivotal question is whether initial impressions introduce “noise” in AC judgments that might blur dimension ratings and validity or whether they contain “substantive” information for predicting performance. For many years, the answer to this key issue has been that initial impressions are inaccurate due to various interpersonal perception errors (e.g., fundamental attribution error, similar-to-me bias, Gray, 2008). This is also the prevailing perspective in ACs. It is argued that intuitive judgments, and thus initial impressions (e.g., on the basis of assessee’s appearance or likability), contain criterion-irrelevant variance and that training should ensure that such errors, biases, and stereotypes do not slip into assessors’ dimension ratings (Thornton & Rupp, 2006).

Whereas past dual process research also highlighted the errors in intuitive judgments, recent attention has shifted toward their adaptive potential. In his Nobel prize lecture, Kahneman (2003) posited that early intuitive judgments serve not only as an anchor for systematic ratings (see H1), but that “intuitive thinking can also be powerful and accurate.” (p. 699). Recently, Evans and Stanovich (2013) have expressed similar views: “In fact, Type 1 processing can lead to right answers” (p. 229). Research on zero-acquaintance judgments confirms this more positive picture; these judgments predicted outcomes as diverse as job performance, teaching performance, music contest outcomes, career success, and firm financial performance (Ambady & Rosenthal, 1992; Rule & Ambady, 2008a; Tsay, 2014). Finally, interview research shows initial impressions predicted invites to another interview and internship offers (Barrick et al., 2010; Swider et al., 2016).

In sum, according to AC textbooks, initial impressions may be prone to biases and harm dimension ratings and criterion-related validity. Yet, evidence that supports these assumptions does not exist. This is an important missing link in AC knowledge because modern dual process theorizing, zero-acquaintance research, and interview studies suggest the opposite: Initial impressions may reflect more than biases and may include valid variance. Given these opposing perspectives, we pose the following research question:

*RQ2:* Do initial impressions contain valid information (vs. biasing information)?

## Method

### Overview

This study obtained data from independent sources: assessors’ AC ratings, raters’ initial impressions, and supervisors’ job performance ratings. Data collection was in line with the ethical standards and approval process of the research institution where the study was conducted. It started with an AC in which assessors provided AC ratings of assesseees and with an online survey of assesseees’ supervisors. The AC data were part of a research project and parts of them have been previously published in an article on dimension transparency (Ingold, Kleinmann, König, & Melchers, 2016). For the present study, only data from the nontransparent AC

condition were used. Afterwards, for the purpose of the present study only, we collected data in two further data collections, during which two rater samples provided initial impressions of these assesseees by reviewing extracted 2-min video clips of their AC performance.

### Actual Assessment Center

**Procedure.** The simulated selection process included two leaderless group discussions and two presentation exercises. We chose these two types of AC exercises because they are the two most commonly used in Western Europe (Krause & Thornton, 2009). The four exercises were designed to simulate a selection process for a management trainee position for graduates and had been successfully used in other studies (e.g., Jansen et al., 2013; Wirz, Melchers, Schultheiss, & Kleinmann, 2014). The AC was designed according to best practices (see Appendix A). There were two administrative (analytical skills, and organizing and planning) and three relational dimensions (consideration of others, persuasiveness, and presentation skills; see Appendices B and C).

**Assesseees.** One hundred and three assesseees participated in the AC as part of a job application training to receive feedback on their performance. Prior to the AC, assesseees agreed to be videotaped for research purposes, provided us with contact details of their current supervisors to collect criterion data and filled in a personality self-report. We excluded three assesseees because their supervisor reported having too few opportunities to evaluate them. The final sample included 100 assesseees (58% male; mean age = 28.76 years,  $SD = 5.71$ ). On average, they had been employed for 2.42 years and worked 28 hr a week. Assesseees worked in diverse sectors: research/education (39%), industrial (10%), service (10%), banking/insurance (8%), media/communication (7%), administration (7%), health/social services (6%), traffic/transportation (3%), sales/distribution (3%), and other sectors (2%).

**Assessors.** Assessors were 35 graduate psychology students (mean age = 28.4 years;  $SD = 8.21$ ) and had participated in a 1-day frame-of-reference training (Roch, Woehr, Mishra, & Kieszczyńska, 2012). To adhere to common practice and to reduce carry-over effects, assessors rotated across exercises (i.e., the same assessors did not rate the same assessee in all exercises).

### Initial Impression Rating Sessions

**Rationale.** Initial impressions were provided by raters that had not participated in the actual AC to avoid demand effects. If we had asked assessors to provide both initial impressions and AC ratings, their initial impressions might have influenced their AC ratings, thereby threatening the interpretation of the results (see Barrick et al., 2010; Lance, Foster, et al., 2004).

**Procedure.** To operationalize the initial impressions sessions, we drew on the zero-acquaintance paradigm (Albright, Kenny, & Malloy, 1988; Borkenau & Liebler, 1992) and collected initial impressions via video excerpts. Consistent with this paradigm and prior research (Barrick et al., 2010; Swider et al., 2016), we extracted the first two minutes per exercise and asked five raters to provide initial impressions of assesseees’ short performance in the exercises. Some videos were excluded due to recording issues. For each video clip, different raters (i.e., a set of five raters in each of the exercises) provided initial impressions of the assesseees. Each

rater saw each assessee only once in one exercise. Afterward, they also rated assessee's appearance and their perceptions of liking. In addition to these ratings, one sample of initial impression raters (see below) also provided their personality impressions of the assessee.

**Rater samples.** We conducted the initial impression sessions with two samples of raters (with assessor experience vs. no experience) for two reasons. First, initial impressions can be influenced by raters' expertise with the social context being assessed (Ambady & Rule, 2007). Thus, we included raters with differing levels of assessor experience to increase the generalizability of our results. Second, using data from two samples avoided common source variance in our analyses (see below). Sample 1 (60% female; age  $M = 32.48$  years,  $SD = 9.11$ ) was recruited from a pool with experience as assessors. On average, they had worked for 3.9 years in HR and had assessed 52 assessee's within the last 3 years. Given their tighter work schedules, Sample 1 did not rate assessee's personality. Sample 2 consisted of psychology bachelor students from a university in Europe (56% female; age  $M = 23.44$  years;  $SD = 6.89$ ).

## Measures

**AC dimension ratings.** Trained assessors rated participants' performance on a 5-point scale (1 = *poor performance* to 5 = *excellent performance*) per dimension in the exercises (see Appendixes B and C). All rating materials had been pilot tested and used in prior studies (Jansen et al., 2013; Wirz et al., 2014). Upon completion of all exercises, assessors discussed their ratings and made adjustments if needed. We computed overall dimension ratings across exercises. Across dimensions, ICCs for postdiscussion ratings ranged between .78 and .88 for a single assessor, and between .88 and .93 for assessors' averaged rating. The criterion-related validity of the dimension ratings ranged from .14 to .29 (see Table 2), which is similar to uncorrected meta-analytic validities (Meriac, Hoffman, Woehr, & Fleisher, 2008). Criterion-related validity of overall AC performance was .24,  $p = .02$ , and thus also similar to uncorrected meta-analytic validities (e.g., Gaugler, Rosenthal, Thornton, & Bentson, 1987; Hermelin, Lievens, & Robertson, 2007; Sackett, Shewach, & Keiser, 2017).

**Personality self-reports.** Assessee's filled in scales for Conscientiousness, Emotional Stability, and Extraversion of the NEO-FFI (McCrae & Costa, 2004) online prior to the AC to receive feedback later. Internal consistencies were .83, .88, and .80.

**Initial impressions.** Eight items measured assessors' general impressions of the 2-min video clip of the assessee, with response options from 1 = *strongly disagree* to 7 = *strongly agree*. An example item is "I have a positive impression of the applicant". Internal consistency was .96 in Sample 1 (.98 in Sample 2). The reliability of a single rating across the exercises (ICC 1.1) was .15 in Sample 1 (.27 in Sample 2). Average ICC (1.5) across exercises was .47 in Sample 1 (.64 in Sample 2). In line with Barrick et al. (2012), we computed a mean initial impression rating across the five raters.

**Appearance ratings.** We measured physical appearance and professional appearance with four items (e.g., "The applicant is attractive"). Response options ranged from 1 = *strongly disagree* to 7 = *strongly agree*. Internal consistency was .74 in Sample 1 (.81 in Sample 2). Average ICC (1.1) across exercises was .23 in

Sample 1 (.27 in Sample 2) and average ICC (1.5) was .59 in Sample 1 (.65 in Sample 2). We computed an overall appearance score across raters.

**Liking ratings.** The scale for liking (Barrick et al., 2010) consisted of three items, with response options from 1 = *strongly disagree* to 7 = *strongly agree*. An example item is "I would like to work with this applicant." Internal consistency was .92 in Sample 1 (.95 in Sample 2). Average ICC (1.1) across exercises was .14 in Sample 1 (.17 in Sample 2) and average ICC (1.5) was .44 in Sample 1 (.49 in Sample 2). We computed an overall liking score across raters.

**Personality impression ratings.** Initial impression raters (only Sample 2) completed a Big Five measure with 30 items in a semantic differential format to assess impressions of assessee's expressed personality (Schallberger & Venetz, 1999; see also Schmid Mast, Bangerter, Bulliard, & Aerni, 2011). We measured all Big Five traits to allow more comprehensive analyses. Internal consistency for Agreeableness, Conscientiousness, Emotional Stability, Extraversion, and Openness to Experience ratings was .93, .94, .94, .93, and .88, respectively. ICC (1.1) ranged from .10 for Openness to .36 for Extraversion and ICC (1.5) from .35 to .74. We computed an average score for each trait across raters.

**Job performance.** Assessee's supervisors were asked to evaluate task performance on a scale from 1 = *not at all* to 7 = *absolutely* via five items from Bott, Svyantek, Goodman, and Bernal (2003). Internal consistency was .83. Supervisors had a mean age of 45.39 years ( $SD = 9.45$ ) and 69% were male. To be included, they were required to have supervised the assessee for at least half a year. On average, they had been in a supervisory position for 4.77 years ( $SD = 1.42$ ).

## Results

H1 posited initial impressions of assessee's to relate positively to AC dimension ratings. We tested this hypothesis by examining the paths from initial impressions (rated by Sample 1) to AC dimension ratings rated by assessors in a path model (Model 1, see Figure 1), using data from different sources to avoid same source bias. All regression weights for the prediction of all five dimension ratings from initial impressions were significant, ranging between .36 for analytical skills and .49 for persuasiveness (see Table 1). This supports H1.

RQ1 dealt with differences in how initial impressions were related to relational versus administrative dimensions. We addressed RQ1 by comparing the fit of two nested path models (Model 2 and Model 3 as variations of Model 1). In Model 2, the path coefficients for the two administrative dimensions and the three interpersonal dimensions respectively were set to be equal, whereas Model 3 constrained all paths from initial impressions to dimension ratings to be equal. The  $\chi^2$ -difference test showed no decrease in fit for Model 3 in comparison to Model 2 ( $\Delta\chi^2 = .001$ ,  $\Delta df = 1$ ,  $p = .97$ ), meaning that Model 3 with equal weights for all dimensions can be kept. Second, a comparison between Model 1 with freely estimated parameters for the path from initial impressions to dimension ratings and Model 3 with equality constraints for all of these paths showed no decrease in fit either ( $\Delta\chi^2 = .492$ ,  $\Delta df = 4$ ,  $p = .97$ ), meaning that Model 3 with equal weights can be kept. Thus, there was no support for initial impres-

Table 1  
Path Coefficients and Indirect Effects of Tested Model

Prediction	Path coefficients to initial impressions			Prediction	Path coefficients to each AC dimension score			Prediction	Path coefficients to job performance		
	$\beta$	$b$ (SE)	$p$		$\beta$	$b$ (SE)	$p$		$\beta$	$b$ (SE)	$p$
Emotional Stability → II	.30	.23 (.09)	.004	II → Analytical Skills	.36	.71 (.20)	<.001	Analytical Skills → JP	.28	.33 (.14)	.022
Conscientiousness → II	.26	.30 (.08)	<.001	II → Organizing & Plan.	.42	.79 (.19)	<.001	Organizing & Plan. → JP	.06	.07 (.19)	.713
Extraversion → II	.20	.15 (.08)	.086	II → Persuasion	.49	.84 (.18)	<.001	Persuasion → JP	-.07	-.09 (.22)	.683
Openness → II	.06	.08 (.13)	.561	II → Cooperation	.41	.71 (.18)	<.001	Cooperation → JP	.04	.05 (.17)	.781
Agreeableness → II	.21	.27 (.15)	.075	II → Presentation Skills	.41	.72 (.16)	<.001	Presentation Skills → JP	.05	.06 (.15)	.663
Appearance → II	.09	.06 (.08)	.428								
Liking → II	.15	.11 (.13)	.415								

Indirect effects of Personality, Appearance, and Liking on Job Performance via AC Dimension Scores

Indirect effect of	Estimate	BC 95% confidence interval
Emotional Stability	.07	.003; .190
Conscientiousness	.09	.004; .237
Extraversion	.04	-.002; .164
Openness	.02	-.042; .149
Agreeableness	.08	-.003; .271
Appearance	.02	-.020; .099
Liking	.03	-.029; .177

Note.  $N = 89$ . II = Initial impressions; JP = Job Performance; BC = Bias-corrected; Personality, liking, and appearance were rated by Sample 2 (without assessor experience), initial impressions by Sample 1 (with assessor experience), AC dimension ratings by the actual assessors, job performance by supervisors. Mediation results displayed are the unstandardized indirect effects and their bias-corrected confidence intervals; 10 000 bootstrap samples.

sions affecting administrative versus interpersonal dimensions differently.

Hypotheses 2, 3, and 4 were also tested via our model, making use of ratings from the two rater samples to avoid same source bias. The regression weights of appearance ( $\beta = .09$ ) and liking ( $\beta = .15$ ) were not significant, hence there was no support for H2 and H3 (see also Table 1). Results supported H4a and H4b because Conscientiousness ( $\beta = .26$ ) and Emotional Stability ( $\beta = .30$ ) significantly predicted initial impressions (see also Table 1). Agreeableness, Extraversion, and Openness did not relate significantly to initial impressions.

RQ2 addressed whether initial impressions contain valid, criterion-relevant information versus errors and biases. We conducted a variety of analyses to answer RQ2. First, we examined whether the relationship between dimension ratings and job performance changed when controlling for initial impressions. If initial impressions introduce error into dimension ratings, controlling for them should increase criterion-related validity. However, if initial impressions capture job-relevant information, controlling for them should lower criterion-related validity. We therefore calculated the correlation between dimension ratings and job performance when partialing out raters' initial impressions in both samples (see Table 3). Compared to the prior zero-order correlations, the correlations decreased for four out of five dimensions in Sample 1 and 2. For analytical skills, there was no change, meaning that initial impressions did not influence this dimension's validity in any direction. Note that the differences in correlations (zero-order vs. partial correlations) were not significant for any dimensions in both samples,  $z_s \leq .40$ ,  $p_s \geq .69$ . These results show that controlling for initial impressions does not increase the validity of dimension ratings. Instead, validities remain constant (for analytical skills) or show a slightly decreasing trend (for four

dimensions). This indicates that initial impressions capture some valid information (instead of criterion-irrelevant information).

Second, we tested the indirect path from personality-based impressions via initial impressions and dimension ratings to job performance. This serial mediation model allows examining whether the personality-based impressions carry job-relevant information via initial impressions to AC ratings that contributes to the prediction of job performance. We used the bootstrapping approach (Preacher & Hayes, 2008) because the sampling distribution of indirect effects is only normal in large samples (MacKinnon, Lockwood, & Williams, 2004; Shrout & Bolger, 2002). Conscientiousness and Emotional Stability exerted an indirect effect on job performance because the confidence intervals did not contain zero (see Table 1). The indirect effects of the other three Big Five, Liking and Appearance were not significant (see Table 1). Thus, personality impressions of Conscientiousness and Emotional Stability carry job-relevant information via initial impressions to AC performance ratings which indirectly contribute to job performance predictions.

Third, we compared effect sizes associated with gender differences for initial impressions and dimension ratings. Differences in initial impression ratings were associated with small to medium effect sizes, with lower ratings for women ( $d = -.19$  in Sample 1;  $d = -.31$  in Sample 2). These  $d$ s did not exceed the  $d$ s for the dimension ratings that ranged from  $-.17$  to  $-.66$ .

Lastly, we examined the correlations between assessors' impressions of assessee's expressed personality and assessee's self-reports on these traits (available for three traits, see Measures). The sizes of the correlations (Conscientiousness  $r = .14$ , Emotional Stability  $r = .32^*$ , and Extraversion  $r = .18$ ) were comparable to the meta-analytic estimates of Connelly and Ones (2010), suggesting these personality impressions can contain some accurate information.

**Table 2**  
*Descriptive Statistics and Intercorrelations of Study Variables*

	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Job performance rated by supervisors	5.73	.87																		
2. Overall initial impression rated by Sample 1	4.13	.36	.15																	
3. Overall initial impression rated by Sample 2	4.83	.51	.14	.79**																
4. Agreeableness rated by Sample 2	4.38	.30	-.12	.20*	.10															
5. Conscientiousness rated by Sample 2	4.50	.33	.06	.41**	.49**	.10														
6. Emotional Stability rated by Sample 2	4.21	.48	.22*	.50**	.68**	-.36*	.14													
7. Extraversion rated by Sample 2	4.03	.50	.19	.56**	.66**	-.03	.07	.64**												
8. Openness rated by Sample 2	3.89	.30	.10	.54**	.39**	.16	.37**	.22*	.60**											
9. Appearance rated by Sample 1	5.18	.45	.07	.39**	.43**	-.15	.24**	.12	.31**	.12										
10. Appearance rated by Sample 2	5.00	.52	.09	.44**	.59**	-.05	.23*	.32**	.42**	.30**	.86**									
11. Liking rated by Sample 1	4.40	.43	.08	.78**	.60**	.25*	.23*	.32**	.51**	.62**	.41**	.49**								
12. Liking rated by Sample 2	4.46	.50	.05	.67**	.81**	.43**	.29*	.45**	.62**	.71**	.42**	.61**	.77**							
13. Analytical skills score rated by AC assessors	3.61	.75	.29**	.37**	.34**	.07	.15	.18	.23*	.08	.09	.11	.26**	.25*						
14. Organizing and planning score rated by AC assessors	3.42	.73	.19	.38**	.45**	-.09	.18	.44**	.41**	.19	.01	.03	.14	.24*	.54**					
15. Cooperativeness score rated by AC assessors	3.74	.64	.17	.35**	.37**	.24	.10	.22*	.32**	.29**	.02	.13	.36**	.38**	.39**	.43**				
16. Persuasiveness score rated by AC assessors	3.65	.70	.14	.40**	.50**	.03	.24*	.39**	.40**	.34**	-.06	.03	.22*	.36**	.30**	.57**	.32**			
17. Persuasiveness score rated by AC assessors	3.59	.64	.18	.46**	.57**	-.10	.28**	.45**	.53**	.28**	.15	.19	.25*	.33**	.61**	.74**	.47**	.47**		
18. Age	28.76	5.71	.11	-.04	.06	-.09	.15	.14	.01	.08	-.10	-.09	-.12	-.02	.02	-.07	-.07	-.07	.05	
19. Sex	1.42	.50	-.07	-.10	-.15	.25*	.16	-.51*	-.05	.10	.22*	.23*	.12	.06	-.09	-.31*	-.16	-.12	-.16	.24*

Note. Sample 1 = raters with assessor experience; Sample 2 = raters without assessor experience; Sex: 1 = male; 2 = female.

\*  $p < .05$ . \*\*  $p < .01$ .

## Discussion

This article set out to advance our understanding of the AC judgment process by examining initial impressions in conjunction with AC dimension ratings for predicting performance. As a first key conclusion, assessors' initial impressions displayed a considerable, positive relationship with dimension ratings. Thus, they seem to serve as an anchor for these subsequent dimension ratings. Initial impressions were also similarly related to all dimension ratings, which helps explain why these ratings are often undifferentiated within an exercise. Note further that initial impressions were not always significantly correlated across exercises (see Appendix D). Thus, they exhibited the same "exercise effects" that have typically been found in analyses of dimension ratings (e.g., Jackson et al., 2016; Lance, Lambert, Gewin, Lievens, & Conway, 2004). This evidence adds important knowledge to the AC domain because it highlights that initial impressions are key drivers behind Lance et al.'s (2004) general impression model underlying dimension ratings.

Second, various analyses suggest that assessors' initial impressions captured some accurate and valid information and did not appear to open up the door for biases and stereotypes. This is because the criterion-related validities of dimensions did not increase when controlling for initial impressions, and the mediation model indicates that the personality-related information captured by initial impressions is criterion-relevant. We found evidence that in the first two minutes assessors pick up personality traits (Conscientiousness, Emotional Stability<sup>2</sup>) that assesseees expressed and that converged with their personality self-reports. Although appearance and liking judgments are quickly available at the start of an exercise, these two aspects did not influence assessors' initial impressions. Moreover, initial impressions did not exhibit larger gender differences than AC dimension ratings. These more positive effects of initial impressions run contrary to common thinking in ACs (cf. Thornton & Rupp, 2006). Hence, further research is needed to build on these findings and examine under which conditions initial impressions relate to AC dimension ratings and validity. We need to examine which other behaviors (e.g., impression management) and attributes might inform assessors' initial impressions. Future studies should also extend our investigation of subgroup differences in initial impression ratings and address to what extent these ratings show ethnic differences (Bobko & Roth, 2013).

Third, our conclusion that initial impression ratings reflect some accurate information should be qualified because it is based on the average initial impression ratings of five raters (see Eisenkraft (2013)). Thus, it should be acknowledged that an *individual* initial impression rating also captures rater idiosyncrasies and shows low interrater reliability (for similar findings in the interview, see Barrick et al. (2012)). One implication is that multiple raters are needed for initial impressions to show acceptable interrater reliability. An intriguing avenue for future studies lies in examining

<sup>2</sup> These traits are not exactly the same as for interviewers' impressions (Barrick et al., 2012). As noted by an anonymous reviewer, in ACs initial impressions develop from scoreable exercises, whereas they rely on rapport building in interviews.



Table 3

*Zero-Order Correlations of AC Dimension Ratings and Task Performance and Partial Correlations of AC Dimension Ratings and Task Performance When Controlling for Initial Impressions*

	Zero-order correlation of dimension rating and task performance <i>r</i>	Partial correlation controlling for initial impressions (Sample 1) <i>r</i>	Partial correlation tested against zero-order correlation (Sample 1) <i>z</i>	Partial correlation controlling for initial impressions (Sample 2) <i>r</i>	Partial correlation tested against zero-order correlation (Sample 2) <i>z</i>
Analytical skills score	.26*	.26*	.00	.27*	-.07
Organizing and planning score	.19	.14	.33	.15	.27
Cooperativeness score	.18	.12	.40	.13	.33
Presentation score	.14	.09	.33	.08	.39
Persuasiveness score	.17	.13	.26	.13	.26

*Note.*  $N = 86$ . Listwise deletion for zero-order correlations and partial correlations.

\*  $p < .05$ .

the effects of increasing the number of raters versus the length of the sample of assessee behavior on reliability and validity.

Fourth, it was noteworthy that analytical skills emerged as the dimension least related to initial impressions ( $\beta = .36$ ) but with the best validity. One interpretation of these results is that the analytical skills rating showed good validity because it was less affected by potentially biasing initial impressions. If this interpretation were correct, one would expect the correlation of analytical skills and job performance (and all other dimensions) to increase when initial impressions were controlled for. However, this was not the case. Therefore, another interpretation seems more on target; that is, these results lend further support that initial impressions seem especially saturated with personality aspects (and potentially less with cognitive ability). Future research should test this further in a variety of ACs and jobs.

Taken together, these conclusions about the role of initial impressions for AC dimensions and their validity have important implications for AC theory. The fact that initial impressions have a substantial overlap with systematic AC dimension ratings and represent some accurate and valid information highlight the role of fast and intuitive judgments in ACs. They also speak in favor of a dual process theory underlying AC judgments that complements the already established elaborate/slow judgment models. This is a key theoretical extension beyond prior AC research that focused only on the systematic judgment route.

This study is not without limitations. First, our initial impressions were based on the first two minutes of an AC exercise. Future research should test the generalizability for different time spans. We also encourage considering the role of initial impressions across the course of an AC. For instance, how do initial impressions that an assessor forms of an assessee from one exercise affect ratings in later exercises? This is challenging to examine because assessor rotation is typically adopted for reducing carry-over effects (Krause & Thornton, 2009), as in this study. Second, we collected data in a simulated AC setting for participants in diverse jobs, which might have decreased the validity of dimension ratings and initial impression ratings. So, this calls for replication of our results in operational ACs that are developed for one particular job. Third, our hypotheses about personality were based on research about the general desirability of traits in selection and were not tied to particular job demands. Future research should examine whether the role of personality traits in initial impressions differs depending on job type.

At a practical level, this study's conclusions speak to the recent popularity of shorter assessment formats (e.g., 3-min role plays; Brannick, 2008; Byham, 2016). Although initial impression ratings seem to serve as anchor for later dimension ratings and can capture some accurate information, we believe that caution is necessary when utilizing them in practice for decision making due to the inherent rater-specific idiosyncrasies (low reliability).

In conclusion, initial impressions provide relevant insights to better understand the AC judgment process and dimension ratings because they serve as a global anchor for later AC dimension ratings and can contain some accurate and valid information based on expressed personality traits. Therefore, initial impressions deserve more attention in AC theory and research. We encourage researchers to further uncover the conditions that affect the relation of initial impressions with dimension ratings, subgroup differences, and job performance.

## References

- Albright, L., Kenny, D. A., & Malloy, T. E. (1988). Consensus in personality judgments at zero acquaintance. *Journal of Personality and Social Psychology, 55*, 387–395. <http://dx.doi.org/10.1037/0022-3514.55.3.387>
- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111*, 256–274. <http://dx.doi.org/10.1037/0033-2909.111.2.256>
- Ambady, N., & Rule, N. O. (2007). Thin slices of behavior. In R. F. Baumeister & K. D. Vohs (Eds.), *Encyclopedia of Social Psychology* (pp. 990–992). Thousand Oaks, CA: Sage. <http://dx.doi.org/10.4135/9781412956253.n585>
- Arthur, W., 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. <http://dx.doi.org/10.1111/j.1744-6570.2003.tb00146.x>
- Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion, 6*, 269–278. <http://dx.doi.org/10.1037/1528-3542.6.2.269>
- Barrick, M. R., Dustin, S., Giluk, T., Stewart, G., Shaffer, J., & Swider, B. (2012). Candidate characteristics driving initial impressions during rapport building: Implications for employment interview validity. *Journal of Occupational and Organizational Psychology, 85*, 330–352. <http://dx.doi.org/10.1111/j.2044-8325.2011.02036.x>
- Barrick, M. R., Swider, B. W., & Stewart, G. L. (2010). Initial evaluations in the interview: Relationships with subsequent interviewer evaluations and employment offers. *Journal of Applied Psychology, 95*, 1163–1172. <http://dx.doi.org/10.1037/a0019918>

- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment, 14*, 317–335. <http://dx.doi.org/10.1111/j.1468-2389.2006.00354.x>
- Bobko, P., & Roth, P. L. (2013). Reviewing, categorizing, and analyzing the literature on black-white mean differences for predictors of job performance: Verifying some perceptions and updating/correcting others. *Personnel Psychology, 66*, 91–126. <http://dx.doi.org/10.1111/peps.12007>
- Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology, 62*, 645–657. <http://dx.doi.org/10.1037/0022-3514.62.4.645>
- Bott, J. P., Svyantek, D. J., Goodman, S. A., & Bernal, D. S. (2003). Expanding the performance domain: Who says nice guys finish last? *The International Journal of Organizational Analysis, 11*, 137–152. <http://dx.doi.org/10.1108/eb028967>
- Bowler, M. C., & Woehr, D. J. (2006). A meta-analytic evaluation of the impact of dimension and exercise factors on assessment center ratings. *Journal of Applied Psychology, 91*, 1114–1124. <http://dx.doi.org/10.1037/0021-9010.91.5.1114>
- Brannick, M. T. (2008). Back to basics of test construction and scoring. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 131–133. <http://dx.doi.org/10.1111/j.1754-9434.2007.00025.x>
- Byham, W. C. (2016, October). *Assessment Centers for Large Populations*. Paper presented at the International Congress on Assessment Center Methods, Bali, Indonesia.
- Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin, 136*, 1092–1122. <http://dx.doi.org/10.1037/a0021212>
- Christiansen, N. D., Hoffman, B. J., Lievens, F., & Speer, A. B. (2013). Assessment centers and the measurement of personality. In N. D. Christiansen & R. P. Tett (Eds.), *Handbook of personality at work* (S. 477–497). New York, NY: Routledge.
- Dean, M. A., Roth, P. L., & Bobko, P. (2008). Ethnic and gender subgroup differences in assessment center ratings: A meta-analysis. *Journal of Applied Psychology, 93*, 685–691. <http://dx.doi.org/10.1037/0021-9010.93.3.685>
- Eisenkraft, N. (2013). Accurate by way of aggregation: Should you trust your intuition-based first impressions? *Journal of Experimental Social Psychology, 49*, 277–279. <http://dx.doi.org/10.1016/j.jesp.2012.11.005>
- Evans, J. S. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255–278. <http://dx.doi.org/10.1146/annurev.psych.59.103006.093629>
- Evans, J. S., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*, 223–241. <http://dx.doi.org/10.1177/1745691612460685>
- Fiske, S. T., & Taylor, S. E. (Eds.). (1991). *Social Cognition* (2nd ed.). New York, NY: McGraw-Hill.
- Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception*. San Diego, CA: Academic Press.
- Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology, 72*, 493–511. <http://dx.doi.org/10.1037/0021-9010.72.3.493>
- Gray, H. M. (2008). To what extent, and under what conditions, are first impressions valid. In N. Ambady & J. J. Skowronski (Eds.), *First Impressions* (pp. 106–128). New York, NY: Guilford Press.
- Hartley, A., Furr, R. M., Helzer, E. G., Jayawickreme, E., Velasquez, K. R., & Fleeson, W. (2016). Mortality's centrality to liking, respecting, and understanding others. *Social Psychological and Personality Science, 7*, 648–657. <http://dx.doi.org/10.1177/1948550616655359>
- Heimann, A. L., Ingold, P. V., Kleinmann, M., Lievens, F., & Melchers, K. (2017, April). *Can we assess the Big Five personality as AC dimensions?* Paper presented at the 32nd Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, USA.
- Hermelin, E., Lievens, F., & Robertson, I. T. (2007). The validity of assessment centres for the prediction of supervisory performance ratings: A meta-analysis. *International Journal of Selection and Assessment, 15*, 405–411. <http://dx.doi.org/10.1111/j.1468-2389.2007.00399.x>
- Highhouse, S. (1997). Understanding and improving job-finalist choice: The relevance of behavioral decision research. *Human Resource Management Review, 7*, 449–470. [http://dx.doi.org/10.1016/S1053-4822\(97\)90029-2](http://dx.doi.org/10.1016/S1053-4822(97)90029-2)
- Human, J. L., & Biesanz, J. C. (2013). Targeting the good target: An integrative review of the characteristics and consequences of being accurately perceived. *Personality and Social Psychology Review, 17*, 248–272. <http://dx.doi.org/10.1177/1088868313495593>
- Ingold, P. V., Kleinmann, M., König, C. J., & Melchers, K. G. (2016). Transparency of assessment centers: Lower criterion-related validity but greater opportunity to perform? *Personnel Psychology, 69*, 467–497. <http://dx.doi.org/10.1111/peps.12105>
- Jackson, D. J. R., Michaelides, G., Dewberry, C., & Kim, Y.-J. (2016). Everything that you have ever been told about assessment center ratings is confounded. *Journal of Applied Psychology, 101*, 976–994. Advance online publication. <http://dx.doi.org/10.1037/apl0000102>
- Jansen, A., Melchers, K. G., Lievens, F., Kleinmann, M., 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. <http://dx.doi.org/10.1037/a0031257>
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*, 697–720. <http://dx.doi.org/10.1037/0003-066X.58.9.697>
- Kahneman, D. (2011). *Thinking, Fast and Slow*. London, England: Penguin.
- Kenny, D. A. (2004). PERSON: A general model of interpersonal perception. *Personality & Social Psychology Review, 8*, 256–280. [http://dx.doi.org/10.1207/s15327957pspr0803\\_3](http://dx.doi.org/10.1207/s15327957pspr0803_3)
- Krause, D. E., & Thornton, G. C., III. (2009). A cross-cultural look at assessment center practices: Survey results from Western Europe and North America. *Applied Psychology, 58*, 557–585. <http://dx.doi.org/10.1111/j.1464-0597.2008.00371.x>
- Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. *Psychological Review, 118*, 97–109. <http://dx.doi.org/10.1037/a0020762>
- Lance, C. E., Foster, M. R., Gentry, W. A., & Thoresen, J. D. (2004). Assessor cognitive processes in an operational assessment center. *Journal of Applied Psychology, 89*, 22–35. <http://dx.doi.org/10.1037/0021-9010.89.1.22>
- Lance, C. E., Lambert, T. A., Gewin, A. G., Lievens, F., & Conway, J. M. (2004). Revised estimates of dimension and exercise variance components in assessment center postexercise dimension ratings. *Journal of Applied Psychology, 89*, 377–385. <http://dx.doi.org/10.1037/0021-9010.89.2.377>
- Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin, 126*, 390–423. <http://dx.doi.org/10.1037/0033-2909.126.3.390>
- MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, 39*, 99–128. [http://dx.doi.org/10.1207/s15327906mbr3901\\_4](http://dx.doi.org/10.1207/s15327906mbr3901_4)
- McCrae, R. R., & Costa, P. T. J., Jr. (2004). A contemplated revision of the NEO five-factor inventory. *Personality and Individual Differences, 36*, 587–596. [http://dx.doi.org/10.1016/S0191-8869\(03\)00118-1](http://dx.doi.org/10.1016/S0191-8869(03)00118-1)

- Meriac, J. P., Hoffman, B. J., & Woehr, D. J. (2014). A conceptual and empirical review of the structure of assessment center dimensions. *Journal of Management, 40*, 1269–1296. <http://dx.doi.org/10.1177/0149206314522299>
- 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. <http://dx.doi.org/10.1037/0021-9010.93.5.1042>
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*, 879–891. <http://dx.doi.org/10.3758/BRM.40.3.879>
- Roch, S. G., Woehr, D. J., Mishra, V., & Kieszczynska, U. (2012). Rater training revisited: An updated meta-analytic review of frame-of-reference training. *Journal of Occupational and Organizational Psychology, 85*, 370–395. <http://dx.doi.org/10.1111/j.2044-8325.2011.02045.x>
- Rule, N. O., & Ambady, N. (2008a). The face of success: Inferences from chief executive officers' appearance predict company profits. *Psychological Science, 19*, 109–111. <http://dx.doi.org/10.1111/j.1467-9280.2008.02054.x>
- Rule, N. O., & Ambady, N. (2008b). First impressions: Peeking at the neural underpinnings. In N. Ambady & J. J. Skowronski (Eds.), *First Impressions* (pp. 35–84). New York, NY: Guilford Press.
- Rupp, D. E., Hoffman, B. J., Bischof, D., Byham, W., Collins, L., Gibbons, A., . . . Thornton, G. (2015). Guidelines and ethical considerations for assessment center operations. *Journal of Management, 41*, 1244–1273. <http://dx.doi.org/10.1177/0149206314567780>
- Sackett, P. R., Shewach, O. R., & Keiser, H. N. (2017). Assessment centers versus cognitive ability tests: Challenging the conventional wisdom on criterion-related validity. *Journal of Applied Psychology, 102*, 1435–1447. <http://dx.doi.org/10.1037/apl0000236>
- Schallberger, U., & Venetz, M. (1999). *Kurzversionen des MRS-Inventars von Ostendorf (1990) zur Erfassung der fünf grossen Persönlichkeitsfaktoren*. Zürich, Switzerland: Psychologisches Institut der Universität Zürich.
- Schmid Mast, M., Bangerter, A., Bulliard, C., & Aerni, G. (2011). How accurate are recruiters' first impressions of applicants in employment interviews? *International Journal of Selection and Assessment, 19*, 198–208. <http://dx.doi.org/10.1111/j.1468-2389.2011.00547.x>
- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*, 422–445. <http://dx.doi.org/10.1037/1082-989X.7.4.422>
- Sutton, A. W., Baldwin, S. P., Wood, L., & Hoffman, B. J. (2013). A meta-analysis of the relationship between rater liking and performance ratings. *Human Performance, 26*, 409–429. <http://dx.doi.org/10.1080/08959285.2013.836523>
- Swider, B. W., Barrick, M. R., & Harris, T. B. (2016). Initial impressions: What they are, what they are not, and how they influence structured interview outcomes. *Journal of Applied Psychology, 101*, 625–638. <http://dx.doi.org/10.1037/apl0000077>
- Thornton, G. C., & Rupp, D. E. (2006). *Assessment centers in human resource management: Strategies for prediction, diagnosis, and development*. Mahwah, NJ: Erlbaum Publishers.
- Thornton, G. C., Mueller-Hanson, R. A., & Rupp, D. E. (2017). *Developing organizational simulations: A guide for practitioners, students, and researchers* (2nd ed.). New York, NY: Routledge.
- Tsay, C. J. (2014). The vision heuristic: Judging music ensembles by sight alone. *Organizational Behavior and Human Decision Processes, 124*, 24–33. <http://dx.doi.org/10.1016/j.obhdp.2013.10.003>
- Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*, 592–598. <http://dx.doi.org/10.1111/j.1467-9280.2006.01750.x>
- Wirz, A., Melchers, K., Schultheiss, S., & Kleinmann, M. (2014). Are improvements in assessment center construct-related validity paralleled by improvements in criterion-related validity: The effects of exercise similarity on assessment center validity. *Journal of Personnel Psychology, 13*, 184–193. <http://dx.doi.org/10.1027/1866-5888/a000115>
- Wojciszke, B., Abele, A. E., & Baryla, W. (2009). Two dimensions of interpersonal attitudes: Liking depends on communion, respect depends on agency. *European Journal of Social Psychology, 39*, 973–990. <http://dx.doi.org/10.1002/ejsp>
- Zedeck, S. (1986). A process analysis of the assessment center method. *Research in Organizational Behavior, 8*, 259–296.

(Appendices follow)

## Appendix A

### Description of the Assessment Center Development

Phases described in Thornton et al. (2017)	Steps taken for developing the AC in this study
Design	Determine purpose: AC for research purposes for graduates with different study backgrounds Desired outcome: An AC that is suitable for the target group and comparable to ACs with regard to dimensions and exercises
Analyze	Simulated context of the AC: Management trainee position for graduates as realistic position for target group
Design	Considering job situations and dimensions that are suitable for the target group Determination of exercises and dimensions in exercises; decision on suitable difficulty level for assesseees
Develop and Pilot	Producing AC material and modifying based on feedback from subject matter experts Piloting with comparable target group
Implement	Actual implementation with recruiting of assesseees, assessor training, revised AC material, time planning, facility booking etc.
Evaluate	Reactions of the assesseees to AC Interrater reliabilities of assessors Criterion-related validity of AC scores

## Appendix B

### Dimension Assessed in the Assessment Center Mapped on Taxonomy of Arthur et al. (2003) and Meriac et al. (2014) and their Frequency of Use

Dimension assessed in the AC	Arthur et al. (2003) dimensions	Meriac et al. (2014) overarching dimension	Percentage of companies in Western Europe reporting use of Arthur et al. (2003) dimensions (Krause & Thornton, 2009)
Analytical skills	Problem solving	Administrative	84%
Organizing and planning	Organizing and planning	Administrative	73%
Persuasiveness	Influencing others	Relational	75%
Cooperation	Consideration and awareness of others	Relational	80%
Presentation skills	Communication	Relational	98%

(Appendices continue)

**Appendix C**  
**Dimension by Exercise Matrix from Assessment Center**

Dimensions	Exercise			
	Presentation 1	Presentation 2	Group discussion 1	Group discussion 2
Analytical skills		x	x	
Organizing and planning	x	x	x	x
Persuasiveness	x	x	x	x
Consideration of others			x	x
Presentation skills	x	x		

**Appendix D**

*Intercorrelations and Descriptive Statistics of AC Ratings and Initial Impression Ratings*

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Overall initial impression rated by Sample 1	4.13	.36															
2. Overall initial impression rated by Sample 2	4.83	.51	.79*														
3. Overall AC performance (AC assessors)	3.59	.55	.49*	.57*													
4. Analytical skills score (AC assessors)	3.61	.75	.37*	.34*	.68*												
5. Organizing and planning score (AC assessors)	3.42	.73	.38*	.45*	.91*	.54*											
6. Cooperativeness score (AC assessors)	3.74	.64	.35*	.36*	.61*	.39*	.43*										
7. Presentation score (AC assessors)	3.65	.70	.40*	.50*	.67*	.30*	.57*	.32*									
8. Persuasiveness score (AC assessors)	3.59	.64	.45*	.57*	.89*	.61*	.74*	.47*	.47*								
9. Initial impression group discussion 1 rated by Sample 1	4.14	.63	.63*	.55*	.24*	.11	.19	.22*	.15	.27*							
10. Initial impression group discussion 2 rated by Sample 1	3.92	.55	.62*	.40*	.22*	.22*	.16	.29*	-.02	.20	.22*						
11. Initial impression presentation 1 rated by Sample 1	4.56	.51	.65*	.53*	.48*	.39*	.32*	.37*	.41*	.44*	.22*	.20					
12. Initial impression presentation 2 rated by Sample 1	3.90	.58	.63*	.52*	.32*	.25*	.28*	.08	.42*	.25*	.11	.13	.29*				
13. Initial impression group discussion 1 rated by Sample 2	4.74	.91	.55*	.66*	.34*	.14	.27*	.30*	.28*	.34*	.70*	.24*	.30*	.13			
14. Initial impression group discussion 2 rated by Sample 2	4.65	.87	.41*	.63*	.35*	.20	.29*	.21*	.13	.39*	.27*	.49*	.14	.15	.20		
15. Initial impression presentation 1 rated by Sample 2	5.22	.68	.45*	.49*	.36*	.28*	.24*	.22*	.46*	.28*	.16	.09	.60*	.31*	.10	.05	
16. Initial impression presentation 2 rated by Sample 2	4.72	.83	.49*	.63*	.34*	.23*	.28*	.14	.37*	.35*	.14	.06	.33*	.68*	.15	.17	.24*

*Note.* Sample 1 = raters with assessor experience; Sample 2 = raters without assessor experience.

\*  $p < .05$ .

Received February 23, 2017  
Revision received May 14, 2018  
Accepted May 14, 2018 ■