1-2012

Core debates in assessment center research: Dimensions 'versus' tasks

Neil D. CHRISTIANSEN

Filip LIEVENS
Singapore Management University, filiplievens@smu.edu.sg

Follow this and additional works at: https://ink.library.smu.edu.sg/lkcsb_research

Part of the Human Resources Management Commons, and the Organizational Behavior and Theory Commons

21

Assessment Centers and the Measurement of Personality

*Neil D. Christiansen, Brian J. Hoffman, Filip Lievens, and Andrew B. Speer***[AQ1]**

Although personality constructs are now widely accepted as being important for understanding work behavior, self-report personality tests as a method of assessment are not without their critics (e.g., Morgeson et al., 2007). Whether misguided or not, concerns persist regarding the validity of these measures and the issue of applicant faking has yet to be fully resolved (Tett & Christiansen, 2007). Moreover, applicant reactions tend to be less favorable for personality inventories than many other assessments commonly used in employment settings (Hausknecht, Day, & Thomas, 2004). To some, self-report inventories may be a poor way to assess personality traits, and yet such inventories are the method most often used to assess these constructs. This trend goes well beyond the area of Industrial and Organizational Psychology. Consistently across the field of psychological measure-ment, personality is rarely formally assessed by directly *observing* a person's behavior (Baumeister, Vohs, & Funder, 2007; for more coverage of personality assessment at work based on observer reports, see Chapter 20, this volume).

An exception to the practice of relying exclusively on the results of paper-and-pencil or com-puterized personality tests is the processes involved in assessment centers (ACs). ACs are composed of multiple moderate to high fidelity simulations of critical job tasks and situations (exercises) where trained raters observe candidates' behavior, and the resulting ratings are used for development or selection purposes. Most often associated with the assessment of managers, the dimensions used to organize the behavioral observations are often aligned to the results of a job analysis or a competency model rather than to any established taxonomy of individual differences. Despite the difference in methods typically used to derive personality traits and AC dimensions, there is obvious overlap in the behavioral domains of each type of construct (e.g., Lievens, De Fruyt, & Van Dam, 2001).

Although personality is not usually explicitly measured in ACs, the relationship between per-sonality and candidate behavior in ACs is intuitive. AC exercises present the opportunity to express a variety of behaviors that are likely a reflection of candidates' personality tendencies (along with abilities and skills). Those individual traits that describe behavior across many other situations should also manifest in an AC. Furthermore, AC assessors are often quick to describe AC candidates in trait terms that any personality psychologist would recognize (Gaugler & Thornton, 1989; Lievens et al., 2001). However, despite the similarities between AC behavior and personality dispositions, social and personality psychologists long ago learned that inferential errors can lead individuals to make dis-positional attributions where trait inferences are at times without foundation (Funder, 1999). Rather than relying on anecdotal observations, theory and research are needed to understand how personal-ity relates to AC processes and outcomes.

Although early views of ACs lacked much theoretical basis, recent developments have linked theories of personality at work to ACs (cf. Haaland & Christiansen, 2002; Lievens, Chasteen, Day, & Christiansen, 2006). These advances promote understanding of the construct relationships of traditional AC ratings and, in particular, how these might be related to aspects of personality. However, just because personality traits may be related to AC behaviors and their resulting dimension ratings does not mean that there is a one-to-one correspondence. The purpose of this chapter is to review both the conceptual underpinnings that relate how personality constructs are assessed in ACs as well as the empirical research that has been done in the area.

AC Method

Borrowing from methods of behavioral simulation developed before and during World War II, Douglas Bray and his team at AT&T developed the first managerial AC. The cornerstone of this approach, like early performance tests, is the focus on behavioral assessment. Physical and psychological fidelity to the criterion domain is viewed as a unique advantage of performance tests relative to other predictors and arguably accounts for the strong psychometric properties of those tests (Wernimont & Campbell, 1968) and favorable user reactions toward them (Hausknecht et al., 2004).

AC exercises have emerged as a distinct subset of performance tests. Although ACs vary widely across administrations and uses (Woehr & Arthur, 2003), they are generally distinguished by some common features. First, ACs are designed to measure behaviors important for effective performance in management and leadership roles. Next, behavioral ratings are provided by trained assessors (Spychalski, Quiñones, Gaugler, & Pohley, 1997). Finally, ACs incorporate multiple simulation exercises designed to elicit behavior relevant to multiple behavioral dimensions. As described below, these core design features have implications for the ways that personality relates to the behaviors displayed in ACs and offer unique potential for the measurement of personality using the AC method.

The behaviors targeted in ACs are most commonly organized into dimensions collapsed across exercises. These dimensions typically reflect skills and competencies that are important for effective performance in leadership roles and the dimension ratings often become the focal scores when interpreting AC performance. Although there is substantial variation in the labels applied to dimensions, there is strong conceptual overlap in the underlying constructs measured in ACs (Arthur, Day, McNelly, & Edens, 2003). Accordingly, researchers and practitioners routinely collapse dimensions into broader, more generalizable "mega-dimensions" when interpreting AC performance (Arthur et al., 2003; Hoffman, Melchers, Blair, Kleinmann, & Ladd, 2011; Shore, Thornton, & Shore, 1990). As we discuss below, these broader dimension factors can be theoretically mapped onto antecedent personality constructs.

AC exercises also differ substantially across administrations. Ideally, AC exercises are designed to reflect a situation common to the focal work role and to elicit behaviors relevant to success or failure. Although exercises were historically viewed as passive vessels by which to measure dimensions, more recent work has aggregated behaviors in ACs based on exercises rather than dimensions, making exercise performance the focal AC constructs (Atkins & Wood, 2002; Collins et al., 2003; Lance, 2008). Similar to dimensions, there is considerable variability in exercises across ACs, but common forms of AC exercises include: leaderless group discussions (LGDs), in-baskets, role-plays, case analyses, and oral presentations. Clearly, these exercises demand different types of behavior for a candidate to be successful in each, making ACs an ideal avenue to measure behavior across unique work situations.

Despite clear agreement on the value of behavioral measurement, the appropriate conceptualization of the behaviors has been controversial in the AC literature. Some argue that the behaviors should be interpreted on the basis of dimensions (Arthur et al., 2003; Arthur, Day, & Woehr, 2008; Rupp, Thornton, & Gibbons, 2008), others point to evidence that the behaviors should be aggregated and

interpreted corresponding to different exercises (Atkins & Wood, 2002; Jackson, Stillman, & Englert, 2010; Lance, 2008; Neidig & Neidig, 1984), and still others take a more multifaceted view of AC performance (e.g., Brannick, 2008; Hoffman et al., 2011; Lievens et al., 2006). Finally, AC performance is often conceptualized using the overall assessment rating (OAR; Gaugler, Rosenthal, Thornton, & Bentson, 1987), which is akin to general performance in the broader performance rating literature. Below we discuss the influence of personality in AC dimensions, exercises, and OARs.

Trait activation theory and ACs

In commercial brochures, it is commonly stated that different AC exercises measure various intrapersonal and interpersonal competencies of candidates. Typical examples are leadership or interpersonal skills. These skill-based constructs are often conceptualized as a reflection of individuals' personality and abilities. In other words, it is assumed that AC exercises allow for the assessment of behavioral expressions caused at least in part by candidates' personality. From a conceptual point of view, one might expect that candidates' behavior shown in AC exercises is related to their standing on personality traits. Trait activation theory (Tett & Burnett, 2003; Tett & Guterman, 2000; Chapter 5, this volume) provides a useful framework to shed light on this personality–AC relationship. In this section, we discuss the basic axioms of trait activation theory. Then, we demonstrate how trait activation theory is relevant to ACs.

As a recent interactionist theory, trait activation theory has foundations in the historical debate in personality and social psychology over the relative importance of traits and situations as sources of behavioral variability. The theory starts with the notion that a person's trait level is expressed as trait-relevant behavior at work. An important underlying principle of the theory is that traits will manifest as expressed work behaviors only when trait-relevant cues are present (Tett & Burnett, 2003). According to trait activation theory, these trait-relevant cues can be categorized into three broad interrelated groups: task, social, and organizational. That is, specific task features (e.g., a messy desk), social features (e.g., problem colleagues), and organizational features (e.g., team-based organizational culture) are posited to influence when and how traits manifest as behavior.

For example, a trait such as autonomy has little opportunity to be expressed in routine monotonous jobs (task level), in the presence of a controlling supervisor (social level), or in a rigid autocratic culture (organizational level), whereas it is more likely to be activated in the opposite conditions. According to trait activation theory, situations are then described on the basis of their situation trait relevance, a qualitative feature of situations that is essentially trait-specific. In essence, it provides information as to which cues are present to elicit behavior for a given latent trait. For example, when an employee is faced with organizing a scattered stack of papers and files on a desk, this situation is relevant for the trait of order (a facet of conscientiousness). Similarly, when someone is confronted with an angry customer, this situation provides cues for traits such as calmness (emotional stability).

A second principle underlying trait activation theory is that trait expression also depends on the strength of the situation (Tett & Burnett, 2003). The notion of situation strength builds on the research about strong and weak situations (Meyer, Dalal, & Bonaccio, 2009; Mischel, 1973). In contrast to situation trait relevance, situation strength is a continuum that refers to how much clarity there is with regard to how the situation is perceived. Strong situations contain unambiguous behavioral demands, where the outcomes of behavior are clearly understood and widely shared. Strong situations and their relatively uniform expectations are therefore likely to result in few differences in how individuals respond to the situation, obscuring individual differences in underlying personality traits even where relevant. Conversely, weak situations are characterized by more ambiguous expectations, enabling much more variability in behavioral responses to be observed. Staying with the same example as above, when a supervisor instructs the employee to clean the messy desk by the end of the shift (with an explicit or implied threat), it will be much more difficult to observe individual

479

differences related to the trait of order, whereas the opposite might be true in the same situation but without clear-cut supervisory instructions.

Thus, according to trait activation theory, the greatest variability in trait-expressive behavior might be observed when individuals act in situations that (a) offer trait-relevant cues (the notion of "situation trait relevance") and (b) are ambiguous (the notion of "situation strength"). Both of these distinct situational characteristics determine a situation's trait activation potential (TAP; Haaland & Christiansen, 2002; Lievens, Tett, & Schleicher, 2009; Tett & Burnett, 2003). So, a situation's TAP is defined as the opportunity to observe differences in trait-related behaviors within a certain situation. The more probable it is to observe these differences, the higher that situation's TAP (see Tett & Burnett, 2003, for a primer on trait activation theory, and the Tett, Simonet, Walser & Brown's chapter in the present volume for an update on the theory).

In ACs, a candidate's rating on a dimension reflects bundles of behavior observed across exercises that may be related to deeper underlying trait or traits. For example, in a role-play with a confrontational supervisor, the candidates may stammer in their verbal responses, resulting in a lower rating on the stress resilience dimension by the assessor. The same behaviors that are targeted by the AC dimension may be expressions of emotional stability. AC exercises, therefore, represent situations that differ in terms of their TAP. The more likely that behavior can be observed within an exercise that is relevant to a particular trait, the higher the exercise's activation potential would be for that trait. The TAP of AC exercises is determined by the availability of trait-relevant cues and the strength of the situation as described above. Apart from the obvious task-demands in the exercise description, advantages of AC exercises over other methods of assessment are the presence of social cues (e.g., clients, colleagues, and supervisors played by role-players or other candidates) and the situation may contain information relevant to a specific organizational culture.

In AC exercises, the strength aspect is represented by the purpose of the AC (a high-stakes selection or promotion opportunity versus a low-stakes developmental purpose) and the specific exercise instructions that provide information and expectations to candidates about what to do to be successful. For example, exercise instructions might mention that the general aim of the exercise is "to reach consensus," "to motivate the problem subordinate," "to make a good impression," or "to give an oral presentation on strategic issues." Of course, instructions and cues regarding effective behavior may come from other sources besides exercise instructions, as candidates could infer what is effective from prior experience with similar situations or from the actions of other candidates.

Taken together, the application of trait activation theory to ACs suggests that the strongest links between candidates' personality trait scores and AC exercises will be found when exercises provide ample cues for behavior related to the trait-related dimensions to be expressed, and when the situations are not too strong to reduce variability in responding to them. To this point, several studies have employed the logic of trait activation theory in ACs by using TAP ratings to determine the cross-exercise convergence of AC dimensions (Haaland & Christiansen, 2002; Lievens et al., 2006). A TAP rating represents the relevance and strength of trait expression in a given situation (Tett & Guterman, 2000). For example, a situation where behaviors related to Agreeableness are expected with some frequency would have a high TAP rating for that trait. Also consistent with trait activation theory, Haaland and Christiansen (2002) found that the relationships between trait scores from a personality test and trait-relevant dimensions were stronger in exercises high in TAP than those evaluated as being low in TAP.

Lievens et al. (2006) recently utilized FFM-based TAP ratings within AC exercises. Experts of both the AC and personality domains linked the FFM traits and a list of common AC exercises. The strongest of these linkages can be seen here in Table 21.1 and represent which trait-relevant behaviors are expected to be easily observed in each AC exercise type. For instance, within the competitive LGD all the FFM traits have high TAP levels and thus behaviors related to these traits may be expected to be easily observed. On the other hand, the role-play has high TAP ratings for

Table 21.1 Linkage Between Five-Factor Model Traits and Typical Assessment Center Exercises

	E	A	C	ES	O
Case analysis			X		
Competitive LGD	X	X	X	X	X
Cooperative LGD	X	X	X		X
In-basket			X		
Oral presentation			X	X	
Roleplay	X	X			

Source: Lievens, Chasteen, Day, and Christiansen (2006).

Note: E: extraversion; A: agreeableness; C: conscientiousness; ES: emotional stability; O: openness to experience; LGD: leaderless group discussion. A cutoff was used to establish trait–exercise linkages.

only extraversion and agreeableness. So, while observations based on these traits might be relatively accurate, inferences regarding conscientiousness, emotional stability, and openness would likely not be so much, as there is less potential to observe trait–related behaviors.

Thus, by using trait activation theory and TAP ratings, researchers and practitioners can determine which personality traits will be relevant for each AC exercise. Likewise, trait activation theory provides a framework for understanding how situational demands affect behavioral expression that may impact ratings on dimensions that have very strong overlap with personality traits. It also informs expectations about the construct–related validity of dimension ratings, as only when two exercises have similar TAP ratings would a high degree of behavioral convergence be expected. This notion has been evidenced by both Haaland and Christiansen (2002) and Lievens et al. (2006), each showing that AC dimension convergence is highest across exercises mutually high in a specific TAP trait and when the dimensions in question are linked to that trait.

Conceptual Overlap Between AC Dimension Ratings and Personality

In ACs, information about behavioral tendencies is evaluated as a method of predicting candidates' behavior in actual work settings, with the behavioral information organized into AC dimensions. At face value, many AC dimensions appear directly related to personality traits. For example, AC dimensions such as sensitivity, drive, and influence would seem relevant to agreeableness, achievement orientation, and extraversion. Despite these similarities, there are also fundamental differences between assessments of AC dimensions and personality traits.

With regard to the overlap between dimensions and traits, evidence shows that, in the process of rating AC dimensions, assessors often describe individuals in terms of personality traits. For example, when taking notes on candidate behaviors, assessors frequently jot down trait inferences and overall impressions rather than actual behaviors (often despite being told explicitly to avoid such attributions). Gaugler and Thornton (1989) demonstrated that 20% to 25% of notes taken by AC assessors contained trait/personality descriptors. A more systematic examination of AC note taking found that, when assessors took notes containing personality–related adjectives, 68% of them could be traced back to the FFM (Lievens et al., 2001). Of these, descriptors related to emotional stability and conscientiousness were mentioned most often. This suggests that assessors organize the behaviors they observe into trait-based schemas when dimension ratings are made.

The Lievens et al.'s (2006) study provides a frame of reference for how AC dimensions map onto personality traits. As part of a study examining the convergence and discrimination of AC dimensions, Lievens et al. (2006) had experts in both the AC and personality domains link Arthur et al.'s (2003) popular taxonomy of seven AC dimensions to the FFM. Table 21.2 displays the strongest linkages

between the FFM traits and AC dimensions. As shown, each FFM trait maps onto at least one of Arthur et al.'s (2003) dimensions, and each of the AC dimensions maps onto at least one of the FFM factors. For instance, there is an intuitive connection between consideration/awareness dimensions and agreeableness, as facets of agreeableness such as tender-mindedness, trust, and altruism (McCrae & John, 1992) overlap with the definition of consideration/awareness (how well one cares for and attends to the feelings and needs of others). Tolerance for stress and uncertainty, defined as "maintains effectiveness in diverse situations under varying degrees of pressure, opposition, and disappointment" (Arthur et al., 2003, p. 136) has obvious links to emotional stability. Problem solving is linked to openness. Extraversion is linked to both communication and influencing others. Conscientiousness is linked to both drive and organization and planning. Thus, despite differences in the constructs typically assessed in ACs and those typically measured with personality measures, there seems to be at least some overlap between the behavioral domains of AC dimensions and personality traits.

Although AC dimensions and personality traits focus on similar sets of behaviors, there are also noticeable differences between the two sets of constructs. First, AC dimensions and personality traits differ in terms of the variety of contexts they are intended to generalize to. Second, the two types of constructs likely differ in how unidimensional (versus multidimensional) each may be. Third, there are differences in the evaluative nature of the constructs in that AC dimensions are value laden whereas the value of trait-related behaviors depends upon situational constraints.

Personality traits are essentially constructs that explain why certain behaviors covary within and across situations. All personality theorists today acknowledge that behavior is a function of both the person and situation and, as such, that behavior will be most consistent within a given situation or across situations with similar demands (Tett & Burnett, 2003). From the perspective of conditional dispositions, trait constructs are better construed as "if-then" propositions that define patterns of behavior in terms that encapsulate both characteristics of people and situations (Mischel & Shoda, 1995). This suggests that trait measures assessed in particular contexts will contain both specific variance and more general variance ultimately attributable to the broad, cross-situational trait constructs.

Consistent with this, research confirms that how extraverted one is at work is related to how extraverted one is in other contexts (Bowling & Burns, 2010; Heller, Ferris, Brown, & Watson, 2009). However, evidence also suggests that prediction of work outcomes is enhanced when personality inventories provide a work frame-of-reference (i.e., with the addition of "at work" tags) and that such measures predict beyond broad assessments (Bing, Whanger, Davison, & VanHook, 2004; Lievens, De Corte, & Schollaert, 2008; Schmit & Ryan, 1993). Thus, there is unique variance involved in contextualized assessment.

Table 21.2 Linkage Between Five-Factor Model Traits and Arthur et al.'s (2003) Assessment Center Dimensions

	E	A	C	ES	O
Communication	X				
Consideration and awareness of others		X			
Drive			X		
Influencing others	X				
Organizing and planning			X		
Problem solving					X
Tolerance for stress and uncertainty				X	

Source: Lievens, Chasteen, Day, and Christiansen (2006).

Note: E: extraversion; A: agreeableness; C: conscientiousness; ES: emotional stability; O: openness to experience. A cutoff was used to establish trait-dimension linkages.

Viewed this way, AC dimensions may be seen as contextualized traits assessed within common work situations, with the ratings of behavioral tendencies not being intended to generalize to other domains of life. In other words, AC designers and raters are not concerned with whether ratings of problem solving predict behavior at a family barbeque, just future behavior on the job as simulated by the given AC exercise (Callinan & Robertson, 2000). In contrast, most personality traits assessed using other methods are broader in terms of context. Thus, while the behaviors that fall under an AC dimension may overlap with aspects of a personality trait, AC dimensions are more narrowly contextualized constructs. This issue of contextualization parallels that of typical and maximal performance (e.g., Sackett, Zedeck, & Fogli, 1988), with ACs constituting maximum performance situations. Due to their short duration and high-stakes nature, ACs are likely to generate behaviors that reflect optimum levels of candidate performance, whereas personality measures typically gauge dispositions that explain candidate tendencies across a range of more common situations. These issues likely have implications for expectations of convergence between AC dimension ratings and noncontextualized personality measures, which generalize to a broader range of situations. As AC dimension scores reflect performance-related behavior, their convergence with personality should occur only to the extent that the traits are also related to performance within the given exercise (i.e., exercise has high TAP).

AC dimensions also differ in that they may not be a function of single individual difference constructs in the way that the behavioral domains of personality traits are generally understood. Similar to many performance dimensions used to evaluate work behavior on the job, individual AC dimensions may represent a hodgepodge of personality traits, abilities, knowledge, and skills. Performance dimensions are typically clusters of activities deemed important by stakeholders, tend to occur in the same venue, or even involve a common piece of equipment; they are not necessarily intended to be activities thought to have one common cause. For example, a dimension involving the retail sales portion of a job may include both dealing with customers and working a cash register, as "warehouse management" for the same position may involve keeping track of inventory levels as well operating a forklift. From that perspective, AC dimensions may be more similar to performance dimensions that index effectiveness as an expression of what the organization values.

Consider an AC dimension such as organizing and planning. Although this dimension could certainly relate to portions of conscientiousness (such as being methodical and industrious), it is also related conceptually and empirically to general mental ability (Meriac, Hoffman, Woehr, & Fleisher, 2008) and might also be linked to facets of openness to experience. To the extent that AC dimensions are more like performance dimensions, ratings will become more akin to composites or indices rather than the scales typically used to measure other predictor constructs. Behavioral convergence would then require the many influences of dimension-related behavior (e.g., knowledge, ability, and personality) to be activated in a similar manner across AC exercises, whereas with a more unidimensional trait (e.g., industriousness), convergence might be easier to obtain.

Personality traits, in contrast, tend to be more unidimensional (at least at the conceptual level), with well-understood behavioral domains. The clarity involved in how a construct is defined and understood by raters can affect both the reliability and validity of ratings. The haphazard way that AC dimensions are often developed and used has been noted in the literature when discussing issues related to construct validity (Arthur et al., 2008; Woehr & Arthur, 2003). For example, dimensions such as inner work standards, personal breadth, inspiring trust, and social flexibility may be commonly used in ACs but are ambiguous in terms of construct domain. As such, the espoused constructs may not be the actual constructs that are being assessed. This may affect how much convergence might be expected between ratings of an AC dimension and any single unidimensional measure.

Finally, there are differences in the evaluative nature of each type of construct. AC ratings are value laden, meaning that high scores on a dimension–effective performance (at least most of the time), whereas trait-relevant behaviors can be viewed either positively or negatively depending on

the situation. Because of this, a trait could be positively related to performance on one exercise but negatively related to performance on another. For instance, assertive behavior might be required in a role-play with an outspoken subordinate, but evaluated negatively in a role-play with an emotionally downtrodden coworker. When an AC dimension would not indicate effective performance in an exercise, it is generally not assessed.

This has implications for expectations regarding the cross-situational convergence of AC ratings. If trait-related behavior is valued in one exercise and not in another, if an AC dimension and personality trait are related and a candidate expresses trait-related behavior similarly across situations, dimension scores may lack convergence (see Haaland & Christiansen, 2002). This becomes more complicated when one considers that (a) multiple personality traits may be related to any single AC dimension; (b) a behavior may be an expression of more than one trait with a given situation; and (c) even if an expression of just one trait, which trait is involved may depend on the context (interrupting may indicate impulsivity or rudeness depending on other behavioral and situational cues). To the extent that assessors are able to take these complexities into account, personality ratings may demonstrate better convergence across exercises that have trait-relevant cues and demands (as trait-related behavior still has implications for trait elevation regardless of whether behavior is regarded as effective or ineffective in that situation). If assessors are not able to take these factors into account due to lack of training or ability, convergence may be worse.

In sum, AC dimensions and personality traits overlap conceptually, but there are notable differences. AC dimensions are work-contextualized characteristics that have a high degree of behavioral overlap with many common personality traits. Although personality traits can also be viewed contextually and assessed in work-specific terms, they are generally conceptualized as having greater situational breadth than traditional AC dimensions. AC dimensions overlap with other skill and ability-based constructs more so than personality traits. They also serve as indices of performance and are designed to overlap more directly with job performance, whereas personality traits can apply to many situations and across many domains. Because AC dimensions act as indices of performance and are thus value-laden, expectations of their cross-situational consistency differ from expectations of trait-related expectations of consistency.

Empirical Relationships Between AC Ratings and Personality

There exists a rich history of studying the relationship between AC ratings and personality. Corresponding with more general trends in the AC literature, this research has operationalized AC performance in a few ways, including: the OAR, dimensional performance, and overall exercise performance. Below, we review evidence pertaining to the relationship between personality and each of the three noted approaches to operationalizing AC performance. Table 21.3 summarizes the results of meta-analytic reviews pertaining to the relationships between the FFM and OARs (Collins et al., 2003; Hoeft & Schuler, 2001), dimensions (Meriac et al., 2008), and exercises (Monahan, Hoffman, Williams, & Lance, 2012).

A first strand of studies focuses on the link between personality and OARs, reflecting the overall clinically or mechanically derived judgment made about candidates at the end of the AC. Similar to measures of job performance criteria (Viswesvaran, Schmidt, & Ones, 2005), AC ratings are characterized by a positive manifold of correlations (Arthur et al., 2003; Hoffman et al., 2011), suggesting a meaningful general factor of performance. Accordingly, the correlations among personality and the OAR give an idea of the personality characteristics that relate to general performance across the tasks and competencies needed for effective performance in the AC.

Three meta-analyses summarize this literature. Collins et al. (2003) meta-analytically investigated the relationship between the OAR and the FFM dimensions. They reported artifact-corrected correlations between $\rho = .17$ and $\rho = .50$ for the personality dimensions of agreeableness, openness,

Table 21.3 Quantitative Summaries of the Relationships Between AC Ratings and Five-Factor Model Personality Domains

	E			A			C			ES			O		
	k	r	ρ	k	r	ρ	k	r	ρ	k	r	ρ	k	r	ρ
Overall assessment rating															
Collins et al. (2003)	13	.36	.50	7	.12	.17				6	.26	.35	5	.18	.25
Hoeft and Schuler (2001)	10	.10	.14	7	−.05	−.07	4	−.05	−.06	8	.12	.15	5	.07	.09
Dimensions (Meriac, Hoffman, Woehr, & Fleisher, 2008)															
Consideration/ awareness of others	9	.07	.10	8	.05	.07	7	.09	.14	4	.07	.10	7	.06	.09
Communication	9	.11	.16	8	.09	.13	6	.09	.12	5	.08	.11	9	.12	.17
Drive	9	.21	.29	6	.09	.12	7	.10	.14	3	.04	.06	7	.06	.08
Influencing others	11	.15	.21	11	.08	.11	6	.09	.13	6	−.01	−.02	10	.08	.11
Organizing and planning	10	.09	.13	9	.02	.03	7	.05	.07	6	.07	.09	10	.09	.12
Problem solving	10	.08	.11	10	.06	.09	6	.13	.17	5	.07	.09	10	.11	.14
Stress tolerance	9	.12	.17	7	.06	.09	3	.12	.17	7	.07	.10	7	.11	.15
Exercises (Monahan, Hoffman, Williams, & Lance, 2012)															
In-basket	7	.06	.07	3	−.02	−.03	4	.13	.16	4	.04	.05	6	.04	.06
LGD	13	.13	.15	10	.00	.00	10	.04	.05	11	.08	.09	10	.07	.08
Roleplay	5	.10	.12	4	.01	.01	5	.02	.02	5	.03	.04	4	.11	.14
Case analysis	3	−.01	−.01	3	−.04	−.05	3	.02	.03	3	.05	.06	2	.12	.15
Oral presentation	3	.13	.17	2	−.10	−.13	3	.09	.11	3	.06	.08	2	.09	.14

Note: E: extraversion; A: agreeableness; C: conscientiousness; ES: emotional stability; O: openness to experience; *r*: sample-size weighted mean observed correlation; *k*: the number of independent samples; r: corrected validity estimate, corrected for attenuation due to predictor and criterion unreliability; LGD: leaderless group discussion.

emotional stability, and extraversion and AC performance, with extraversion as the strongest FFM predictor of OAR. Scholz and Schuler (1993) also conducted a meta-analysis of studies in which AC scores were correlated with an array of external measures such as personality inventories. This meta-analysis revealed that the overall AC rating tended to correlate .23 (corrected for unreliability) with dominance, .30 with achievement motivation, .31 with social competence, and .26 with self-confidence. In another meta-analysis that also examined personality correlates of the OAR (Hoeft & Schuler, 2001), much lower correlations were reported with the personality traits of agreeableness ($\rho = −.07$), conscientiousness ($\rho = −.06$), openness ($\rho = .07$), extraversion ($\rho = .14$), and emotional stability ($\rho = .15$).

Taken together, these studies produced equivocal results. There are four potential explanations for the mixed findings. Most obviously, more narrow personality constructs were used across one of the supportive reviews; perhaps constructs more clearly targeted to the domain will have stronger relationships than broad bandwidth constructs, as in the Scholz and Schuler's (1993) review. Second, each review targeted a somewhat different literature base. Third, some meta-analyses may have included studies wherein the OAR was based on information that came not only from AC exercises but also from personality inventories, artificially inflating observed relationships (Collins et al., 2003). Finally, the use of the OAR has been criticized because it can potentially obscure effects by combining unique aspects of performance across dimensions and exercises (Arthur et al., 2003; Arthur et al.,

2008). Instead, some have argued that theoretically mapping personality onto unique aspects of AC performance can provide a more meaningful analysis of the construct validity of AC ratings. In the next two sections, we discuss the findings of research linking personality to ratings of performance in AC dimensions and exercises.

Personality and AC Dimension Ratings

The correlation between personality and AC dimensions has important implications for understanding the nomological network of AC dimensions (Cronbach & Meehl, 1955; Shore et al., 1990) and the potential for incremental validity of AC dimensions beyond other measures (Meriac et al., 2008). It is important to investigate the construct-related validity of dimensions, given that AC dimensions are often the focus when interpreting AC performance, especially in developmental contexts. Research attempting to support the nomological network of dimensions by using external measures of personality has yielded mixed results.

Although some studies found support for a relationship between AC dimension ratings and conceptually related personality trait ratings (e.g., Dilchert & Ones, 2009; Shore et al., 1990; Thornton, Tziner, Dahan, Clevenger, & Meir, 1997), such hypothesized convergence has not always been confirmed (e.g., Chan, 1996; Fleenor, 1996; Goffin, Rothstein, & Johnston, 1996). In these latter studies, the final dimension ratings failed to demonstrate expected relationships with conceptually similar personality dimensions. Furthermore, the average correlations between final dimension ratings and conceptually dissimilar personality dimensions were equal to or even higher than those with conceptually related personality dimensions.

Meriac et al. (2008) presented a meta-analysis of the relationship between individual differences and final dimension ratings using Arthur et al.'s (2003) seven-dimensional taxonomy as an organizing framework. Their results indicated generally weak and inconsistent relationships between AC dimensions and personality (ρs ranging from $-.11$ to $.29$). There was some evidence for the nomological network of dimensions, with general mental ability correlating more strongly with problem-solving dimensions than with interpersonally oriented dimensions, and with extraversion significantly correlating with influencing others. However, the hypothesized correlations between organizing and planning and conscientiousness were weak in this review. As might be expected based on the inconsistent support in the literature, only modest support was provided for the nomological network of dimensions based on Meriac et al.'s (2008) meta-analysis.

In sum, studies of the overlap between personality constructs and AC dimension ratings show an equivocal picture. The primary studies have come to differing conclusions, and the largest existing meta-analysis provided only modest support. A possible explanation for these findings is that the labels applied to certain AC dimensions do not match the actual construct that is measured (Arthur & Villado, 2008). However, although incorporating theoretical taxonomies of AC dimensions has improved findings in some studies (Dilchert & Ones, 2009; Shore et al., 1990), it has not in other studies (Meriac et al., 2008). It is possible that closer attention needs to be paid to the dimensions underlying broad factors in existing conceptual taxonomies. For instance, Arthur et al.'s (2003) popular seven-dimensional taxonomy is rarely supported empirically; instead, 2–4 dimensions more regularly describe the structure of final dimension ratings (Hoffman & Woehr, 2009; Kolk, Born, & Van der Flier, 2004; Schmitt, 1977; Shore et al., 1990). Thus, perhaps this taxonomy specifies too many dimensions to reasonably expect differential relationship with personality variables. Similarly, there are questionable linkages between the subordinate dimensions assigned to Arthur et al.'s taxonomy. For instance, communication includes written communication, which is strongly correlated with GMA**[AQ2]**, possibly resulting in the larger-than-expected relationship between communication and intelligence and the weaker relationships between communication and personality. Similarly, organizing and planning includes subordinate dimensions of developing others and control, more

486

routinely included under leadership behaviors (cf. Borman & Brush, 1993). This possibly explains why extraversion was the strongest personality predictor of organizing and planning in the Meriac et al.'s (2008) review.

Personality and AC Exercises

Given the well-known measurement problems with AC dimensions (Lance, 2008), a third group of studies examined the relationship between personality and AC exercises by focusing on AC exercise scores instead of AC dimension scores (Craik et al., 2002; Lievens et al., 2001; Spector, Schneider, Vance, & Hezlett, 2000). Although somewhat novel to the AC literature, research stemming from other areas has frequently investigated personality predictors of overall performance in behavioral simulations (Brunell et al., 2008; Foti & Hauenstein, 2007). The central prediction is that the behaviors manifest in a given situation (exercise) will be a reflection of an individual's underlying personality in response to the demands of the situation. Because exercises differ in their general potential to activate behavior related to specific traits, it is expected that the relationship between personality scores and AC exercises will differ, depending on the type of exercise.

Although only a handful of studies have directly examined the overlap between personality and AC exercises (Craik et al., 2002; Spector et al., 2000), a recent meta-analysis presented correlations between individual differences and exercise performance (Monahan et al., 2012). A common finding in existing research is that extraversion is consistently shown to be among the best personality predictors of performance in LGDs (Craik et al., 2002; Monahan et al., 2012). This is consistent with leadership literature supporting extraversion as a key antecedent to leader emergence (Judge, Bono, Ilies, & Gerhardt, 2002) and with Lievens et al.'s (2001) findings that notes on assessor's LGD rating forms were characterized mainly be extraversion descriptors. In addition, intelligence and conscientiousness, two constructs having particularly close ties to the completion of task responsibilities are more strongly related to in-basket performance than to performance in other exercises (Craik et al., 2002; Monahan et al., 2012). Given the task, rather than interpersonally oriented nature of the in-basket responses, this finding supports the nomological network of in-basket scores. This pattern of results is also consistent with Lievens et al.'s (2001) finding that assessors' rating sheets most frequently included conscientiousness descriptors in the in-basket exercise. Thus, some of the observed results support the predictability of exercise performance based on individual differences.

On the other hand, FFM traits were weakly and somewhat sporadically related to role-play, case analysis, and oral presentation exercises. Openness was the strongest personality correlate of role-play performance in the Monahan et al. (2012) review, and this correlation was quite weak ($\rho = .14$). This overlap with openness might reflect the ability to use one's imagination in order to "get into role" in the sometimes awkward role-play simulation. This suggestion is consistent with Meriac et al.'s (2008) results that openness is among the strongest trait predictors of stress tolerance. Finally, limited research has examined individual difference correlates of case analysis exercises and oral presentations; however, there is some evidence that extraversion predicts oral presentation performance (Monahan et al., 2012).

Basically, these studies provide some general support for trait activation theory as they show that personality is differentially related to performance on different AC exercises. However, the support is most pronounced in correlations of in-basket and LGD performance, as the correlations between other exercises and personality seem to be less predictable. These studies, however, do not test whether specific exercise stimuli elicit specific trait-related behavior. They also do not reveal whether interventions to increase the situational trait relevance and strength of AC exercises affect the link between personality and AC ratings.

Moderators of Personality–AC Rating Relationships

A common thread running through all aforementioned strands of research is that they focused on the main effect of personality on AC ratings (either conceptualized at the OAR, dimension, or exercise level). A final stream of studies has searched for moderators of the personality–AC relationship, aiming to explain under what conditions personality might relate to AC ratings (Jansen, Lievens, & Kleinmann, 2011; Kolk et al., 2004; Krajewski, Goffin, Rothstein, & Johnston, 2007).

Some of these studies have built on trait activation principles, aiming to uncover conditions that might trigger or constrain personality trait expression in ACs. In this respect, Jansen et al. (2011) discovered that relevant traits were triggered only when candidates perceived the situation demands correctly. The general hypothesis was that individual differences in people's perception of situational demands will moderate the relationship between personality traits and conceptually related AC dimension ratings because only among candidates who perceived that a given exercise required behavior related to given personality traits would those traits be expressed behaviorally. This logic was confirmed for two of the three traits: agreeableness and conscientiousness. In particular, Jansen and colleagues showed that self-report agreeableness was related to ratings on cooperation in AC exercises only among people who perceived that the situation demanded agreeable behavior. Similar results were obtained for the relationship between participants' standing on conscientiousness and their AC rating on planning and organizing.

Krajewski et al. (2007) argued that age might moderate the relationship between personality and managerial effectiveness as measured by the AC. In particular, they posited that older managers with high scores on certain job-related personality traits may express them in a more effective fashion than similar younger managers, thus causing age to moderate the relationship between personality and AC performance. Consistent with hypotheses, age moderated the relations of dominance and exhibition with AC performance, such that dominance and exhibition were more strongly related to AC performance for older as opposed to younger managers.

Finally, Kolk et al. (2004) did not focus on trait-expression moderators but examined three method-related factors that might moderate the personality–AC relationship, namely differences in rating source (other vs. self), rating domain (general vs. specific), and rating format (multiple items vs. single item). For instance, the hypothesis about rating source was that the correlation between personality and AC ratings would be higher when the rating source was held constant across the personality inventory and the AC. There was partial support for the influence of each of the three method factors, although the differences were not large.

A few noteworthy trends emerged across investigations of the influence of personality on AC ratings. First, the three most recent meta-analyses revealed relatively weak relationships between personality and AC scores, regardless of whether AC performance is operationalized using OAR, dimensional performance, or exercise performance. This weak correlation should not be surprising, given that ACs are rarely used to measure personality; instead, the competencies measured in ACs are thought to be a consequence of individual personality, ability, experience, and skills. Indeed, if ACs overlapped substantially with self-reports of personality, there would be little need to go through the time and expense of administering an AC. It is also noteworthy that the magnitude of correlations between AC ratings and the FFM are relatively consistent with those investigating the FFM and job performance ratings (e.g., Barrick & Mount, 1991).

Although the modest magnitude of the correlations is not problematic for AC ratings, the moderate support for the nomological network of individual differences and AC ratings is more troubling. Specifically, for the most part, the three most recent reviews did not provide particularly strong evidence for differential relations among AC ratings and theoretically related (and unrelated) personality constructs. However, across all three reviews, AC ratings were more strongly related to GMA than to self-reports of personality (Collins et al., 2003; Meriac et al., 2008; Monahan et al., 2012). In

addition, cognitively oriented dimensions and exercises generally correlate more strongly with GMA than do interpersonally oriented dimensions and exercises. The more influential impact of GMA on performance in ACs should not be surprising, given that GMA is found to be among the strongest predictors of performance across settings and performance domains (Schmidt & Hunter, 1998).

Next, extraversion and conscientiousness emerged as the two strongest personality predictors of AC performance across all three AC scoring approaches, and emotional stability and agreeableness were less strongly related to AC performance, though these differences were not large. Still, given the theoretical links between extraversion and leadership performance (Judge et al., 2002) and conscientiousness and performance across settings (Barrick & Mount, 1991), this pattern of results provides modest evidence for the construct-related validity of AC ratings.

Using Behavioral Observations to Directly Measure Personality in Work Simulations

Despite the widespread use of personality assessment today, little work has been done to measure traits directly in work situations such as ACs. This would ease the inferential leap involved in mapping traits onto AC dimension ratings if the results of the two types of assessments are to be combined mechanically. There may also be advantages in that, as constructs, personality traits such as the FFM are well understood in terms of their behavioral domains and place in the nomological network. This may facilitate the development of predictive hypotheses based on past research or even prove beneficial in terms of the psychometric properties of the ratings. Unfortunately, there is currently very little in the way of existing formal, behavioral assessments of personality that are suitable for work situations.

However, recent research has developed a tool (the Work Simulation Personality Rating Scales [WSPRS]) to assess personality-related behavior in work simulations and provides evidence of reliability and validity in an AC context (Christiansen, Honts, & Speer, 2011). The WSPRS is a 40-item measure designed to assess behaviors relevant to the FFM and common to work situations, with eight items per FFM trait. The WSPRS items were developed based on existing AC instruments, behavioral coding schemes (Funder, Furr, & Colvin, 2000), and traditional self-report personality inventories of the FFM (Goldberg et al., 2006). The items were formed at a moderate level of abstraction and specifically for application in a work simulation context. Furthermore, the WSPRS was grounded in trait activation theory (Tett & Burnett, 2003), meaning potential utility of the instrument was assumed to be dependent upon the situation in which behavior was assessed. Only in situations in which trait behavior is relevant and where behavioral variation is expected should the instrument be expected to assess a given FFM trait accurately.

A study was conducted to evaluate the WSPRS by applying it to 123 candidates of a developmental AC consisting of five behavioral exercises (three role-plays, an LGD, and a case analysis presentation). Raters were trained using frame-of-reference procedures on sample videotapes before viewing actual videotapes of candidates within the AC. Three assessors rated each candidate using the WSPRS. Traditional AC ratings were also collected by a separate set of trained assessors, and ratings of TAP were completed as a check on the assumption that behavioral observations would be more accurate in situations deemed trait-relevant. The observer ratings of personality were then correlated with scores on a self-report personality test completed by the candidates prior to the AC.

Results revealed that behavioral observations of the WSPRS reached moderate convergence with self-rated personality scores (see Table 21.4). Uncorrected correlations between self-report and WSPRS dimension scores ranged from a low of .11 for emotional stability to a high of .31 for extraversion. The magnitude of these correlations is similar to those presented by Connelly and Ones (2010) for the correlations between self-ratings and stranger ratings (see Figure 21.1). Interestingly,

Table 21.4 Correlations Between Behavioral Observation Ratings and Self-Reported Personality

	TAP	r	r
Extraversion	3.88	.31	.36
Agreeableness	3.41	.24	.31
Openness	3.02	.22	.30
Conscientiousness	2.94	.18	.22
Emotional stability	2.80	.11	.13

Note: TAP: trait activation potential (these ratings are on a 1–5 scale); IRR: interrater reliability for a composite of three independent raters; r: uncorrected correlation between WSPRS composites and self-report scores; ρ: corrected correlation between WSPRS composites and self-report scores, corrected for self-report unreliability; WSPRS: Work Simulation Personality Rating Scales.
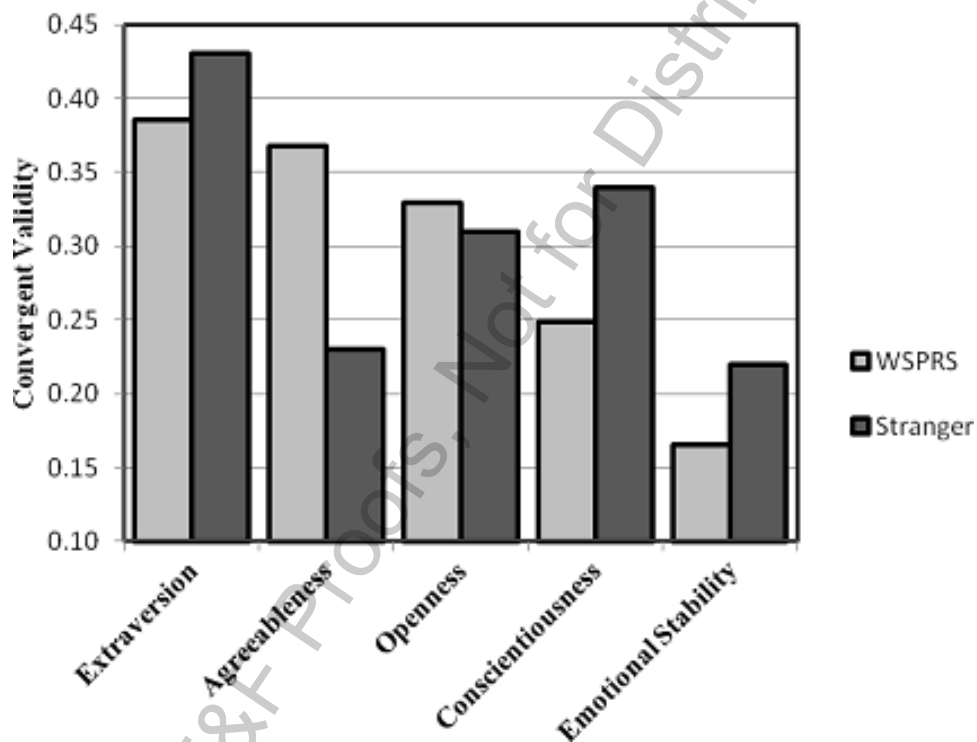
Source: Christiansen, Honts, and Speer (2011).

Figure 21.1   Convergent Validity Estimates From Correlating Self-Report Personality Measures With Observer Ratings of Strangers.

Note: WSPRS: Work Simulation Personality Rating Scales. Estimates for stranger ratings taken from Connelly and Ones (2010). Correlations are corrected for both unreliability in self-report measures and interrater unreliability in the observer ratings.

the degree of self–WSPRS convergence almost directly coincided with the rank ordering of TAP for the FFM traits, in that WSPRS dimension scores correlated more strongly with corresponding self-ratings when there was ample opportunity to observe trait–related behaviors. For example, emotional stability was rated lowest in TAP and likewise had the lowest convergence with self-reports

of that trait. Extraversion was highest in TAP and had the highest correlation between the WSPRS and self-report scores. The same general trend occurred in terms of interrater reliability, where the average interrater reliability of a composite of three raters ranged from .62 to .87 for the FFM dimensions (see Table 21.4). As expected, observations using the WSPRS were more reliable and more accurate for those traits that were relevant to the situation.

Table 21.5 displays a list of WSPRS items and item characteristics. Items such as "exhibits high enthusiasm and energy," "appears passive," "acts in a polite manner toward others," "contributes new and creative ideas," and "attempts to keep group organized" all had appropriate variability and moderate to high relationships with other items in their respective scales. On the other hand, some items had low variation and did not correlate well with other items. For instance, the items, "behaves in a non-normative manner" and "behaves in a rude or abrupt manner" did not discriminate well amongst AC candidates. The items, "argues their opinion or point" and "openly emotional and/or volatile" did not correlate with other items in their respective scales. It is unclear whether these items were poor representatives of their targeted traits or whether the situations did not allow enough opportunity for the expression of these behaviors (i.e., low baseline).

Table 21.5 Work Simulation Personality Rating Scale Item Statistics and Convergence With Self-Report Personality

	Descriptive Statistics		Reliability		Self-Report Convergence	
	M	SD	IRR	CITC	r	ρ
Extraversion						
Behaves in an influential and persuasive manner	1.92	.59	.70	.80	.32	.37
Exhibits high enthusiasm and energy	1.79	.63	.83	.83	.31	.36
Talkative	2.06	.62	.81	.82	.28	.33
Interacts confidently with others	2.23	.59	.77	.75	.28	.32
Expressive with voice, face, and/or gestures	1.97	.63	.75	.79	.26	.30
Seems detached from the situation (R)	2.49	.54	.77	.74	.25	.29
Appears passive (r)	2.34	.63	.80	.80	.24	.27
Behaves timidly (r)	2.44	.56	.75	.78	.21	.24
Agreeableness						
Makes supportive comments	2.07	.58	.54	.70	.27	.35
Expresses agreement or support	2.12	.46	.57	.70	.27	.35
Acts in a polite manner toward others	1.98	.61	.63	.69	.24	.31
Displays concern for others	2.01	.52	.62	.67	.19	.25
Supports others' decisions	2.21	.46	.56	.67	.16	.20
Behaves dismissively toward others (r)	2.56	.46	.57	.66	.12	.15
Behaves in a rude or abrupt manner (r)	2.65	.44	.56	.65	.11	.14
Argues their opinion or point (r)	1.93	.57	.65	.16	.01	.01
Openness to experience						
Says interesting things	1.89	.57	.67	.81	.25	.37
Discusses multiple aspects of ideas and topics	2.01	.52	.55	.80	.25	.36
Considers both pros and cons	1.99	.57	.66	.82	.22	.32
Exhibits a high degree of intelligence	2.02	.58	.72	.88	.18	.26
Integrates others ideas and suggestions	2.11	.52	.55	.74	.17	.25

(Continued)

Table 21.5 (Continued)

	Descriptive Statistics		Reliability		Self-Report Convergence	
	M	SD	IRR	CITC	r	ρ
Contributes new and creative ideas	1.97	.58	.73	.72	.14	.21
Makes nonintellectual statements (r)	2.46	.48	.50	.65	.14	.20
Unconcerned with different thoughts and ideas (r)	2.46	.48	.57	.60	.09	.13
Conscientiousness						
Emphasizes goals and accomplishments	2.07	.58	.72	.74	.22	.27
Attempts to keep group organized	1.88	.61	.70	.77	.17	.21
Encourages group to stay on task	1.87	.57	.67	.80	.16	.20
Does not behave professionally (r)	2.48	.52	.70	.67	.16	.19
Prioritizes or plans activities	2.07	.58	.72	.80	.15	.18
Dresses appropriately	2.46	.56	.76	.43	.13	.15
Easily distracted and does not follow through (r)	2.56	.48	.70	.72	.07	.09
Considers all options and is thorough	2.00	.57	.66	.67	.03	.04
Emotional stability						
Appears calm and relaxed (r)	1.60	.46	.40	.44	.12	.14
Interacts poorly or awkwardly	1.57	.52	.59	.73	.08	.10
Interested in others and tasks (r)	1.77	.52	.68	.48	.08	.10
Behaves in a non-normative manner	1.36	.41	.48	.61	.08	.09
Acts irritated or annoyed	1.37	.42	.55	.25	.08	.09
Openly emotional and/or volatile	1.27	.37	.57	.19	.07	.08
Seeks reassurance from others	1.49	.48	.56	.45	.02	.02
Displays low opinion of self	1.36	.43	.58	.46	.01	.02

Source: Christiansen, Honts, and Speer (2011).

Note: *M*:   mean, with items on a 1–3 scale; *SD*, standard deviation; IRR, interrater reliability for a composite of three independent raters; CITC, corrected item total correlation; *r*, uncorrected correlation between WSPRS items and self-report composite scores; ρ: corrected correlation between WSPRS items and self-report composite scores, corrected for self-report unreliability; (r): an item that was reverse-coded; WSPRS: Work Simulation Personality Rating Scales.

The WSPRS scores were also correlated with the OAR to determine the degree of overlap. Essentially, the OAR represents assessment performance and is often used to make personnel decisions in operational ACs. As both the OAR and WSPRS ratings were based on the same videotaped sets of performance, it was expected that they would be highly correlated, which they were (R = .73). Extraversion emerged as the best predictor of the OAR (β = .34), followed by openness to experience (β = .24), conscientiousness (β = .16), agreeableness (β = .08), and emotional stability (β = .05).

To date, this research represents the most direct evidence that behavior relevant to personality can be readily observed in AC exercises. It also underscores how important a thorough understanding of exercise demands is in terms of how well traits can be assessed. In this AC, emotional stability in particular was hypothesized to be the most difficult to observe relevant behavior and evidence confirmed this. If the AC were being used for a position where this trait is critical, existing exercises would need to be redesigned or additional exercises added to provide more trait-relevant cues. On the other hand, with the current exercises (three role-plays, group discussion, and a presentation), many cues were present for extraversion.

Effects of Impression Management and Response Distortion

Another potential advantage of assessing personality in the context of ACs is that it is likely to be more difficult to raise scores by engaging in impression management. Research has shown that response distortion is relatively common when personality tests are used in applicant settings and degrades their validity (see Tett & Christiansen, 2007). In most self-report inventories, it is fairly easy to identify the response that is favorable for a job (Christiansen, Burns, & Montgomery, 2005). On the other hand, research has shown that faking is more difficult when using interviews to assess personality, where applicants need to generate and describe job-related examples that are plausible, detailed, and relevant to the questions asked (Van Iddekinge, Raymark, & Roth, 2005). In addition, overtly obvious attempts at impression management may result in negative evaluations (see Chapter 18, this volume).

Although no research has directly examined the effects of impression management in AC in terms of mean shifts or validity, it would seem more challenging to raise scores in this context than when an interview or a personality inventory is used to measure personality. To achieve high scores in an AC exercise, candidates must first determine what responses will be most effective both in terms of task success and the perceptions of evaluators. This can be challenging because the situations in ACs are typically much more complicated than a question on a personality test or interview. Beyond just determining a desirable response, in ACs candidates have to actually engage in behavior that results in assessors evaluating them more favorably, rather than just saying they would do it or had done it in the past. This places greater cognitive demands on AC participants, forcing them to focus more on the task at hand and in turn limiting their ability to engage in impression management (McFarland, Ryan, & Kriska, 2003). Similar to interviews, overt attempts at impression management may be expected to result in negative evaluations in ACs, whereas they are seldom taken into account by assessors when reviewing the results of self-report personality inventories (Christiansen, Burns, & Rozek, 2010).

This is not to say that impression management does not happen in ACs. The prevalence and effects are likely to depend on the demands of the exercise where trait–relevant behavior is to be observed. For example, impression management is less common in exercises demanding technical competency than in those requiring interpersonal effectiveness in order to be successful (McFarland, Yun, Harold, Viera, & Moore, 2005). Even so, the effects of impression management on criterion–related validity may be at least partly mitigated, as those who are correctly able to identify what they are being evaluated on are also likely to perform better on the job (Kleinmann et al., 2011).

Conclusions and Future Directions for Research

Modern research on ACs and personality measurement has largely progressed independently, with limited attempts at empirical integration. In this chapter, we have attempted to highlight the similarities and differences between the constructs assessed in ACs and traditional self-report measures of personality, the theoretical and empirical overlap between ACs and personality, and ways in which AC research and practice might inform the assessment of personality. This review shows that, although performance behaviors in ACs is likely a function of personality, the constructs measured in ACs tend to be modestly related to self-reports of personality, regardless of whether AC performance is assessed in terms of dimensions, exercises, or the OAR. Nevertheless, the pattern of observed relationships does provide some support for the construct–related validity of both ACs and self-reports of personality.

Together, although ACs do tap some relevant aspects of personality, the modest correlations between the two suggest that the constructs measured in typical ACs are not interchangeable with those measured with personality scales. That said, it is difficult to know whether correlations are weak

493

because AC dimensions and personality constructs are actually that different or because ACs and self-reports of personality reflect fundamentally unique methods of measurement (see Arthur & Villado, 2008). Future research directly comparing the psychometric soundness of behavioral and self-report based measures of personality could help to clarify this issue. The WSPRS research presented here seems a promising starting point. Similarly, it might be interesting to compare self-reports of traditional AC competencies to assessor ratings. By comparing different methods of measuring the same constructs, it might be possible to enhance the accuracy of prediction associated with these tools.

Practitioner's Window

1. Given the modest relationships that have been observed between measures of personality and AC ratings, practitioners should consider using both trait and AC dimensions as complementary sources of information.
2. Assessment of personality traits in ACs could take on a number of forms, such as:
    *   The traditional approach of administering a personality inventory alongside simulation exercises. This may have merit but trait scores and AC dimension ratings should be mechanically combined into more broad composites reflecting "mega-dimensions." We suspect many ACs have utilized personality test results subjectively when making final dimension ratings, often as part of the consensus process. In either event, traits and dimensions should be conceptually linked by experts familiar with both types of constructs.
    *   An alternative approach that focuses on behavior within the AC (rather than self-report) would be to require assessors to write down adjectives descriptive of personality when observing candidates in simulation exercises. Assessors could then make judgments of relevant traits on rating scales that could be combined with dimension ratings (or not), as above.
    *   Practitioners could formalize this by using specific behavioral scales, such as the WSPRS, that have recently been developed to assess personality-related traits directly in simulation exercises. Initial investigation into the reliability and validity of these scales has been promising.
3. Exercises should be evaluated in terms of trait activation potential and traits to be considered in each winnowed to those where there is ample opportunity to observe trait-relevant behavior. For that matter, practitioners should consider carefully which traditional AC dimensions can reasonably be assessed in each exercise in order to further reduce the cognitive load on the assessors. To the extent that a dimension has strong links to personality traits, it may be difficult to rate candidates accurately if behavior related to those traits is judged by experts to have little opportunity to be observed.

References

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.

Arthur, W., Jr., Day, E. A., & Woehr, D. J. (2008). Mend it, don't end it: An alternate view of assessment center construct-related validity evidence. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 105–111.

Arthur, W., & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology, 93*, 435–442.

Atkins, P. W., & Wood, R. E. (2002). Self- versus others' ratings as predictors of assessment center ratings: Validation evidence for 360-degree feedback programs. *Personnel Psychology, 55*, 871–904.

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

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

Bing, M. N., Whanger, J. C., Davison, H. K., & VanHook, J. B. (2004). Incremental validity of the frame-of-reference effect in personality scale scores: A replication and extension. *Journal of Applied Psychology, 89*, 150–157.

Borman, W. C., & Brush, D. H. (1993). More progress toward a taxonomy of managerial performance requirements. *Human Performance, 6*, 1–21.

Bowling, N. A., & Burns, G. N. (2010). A comparison of work-specific and general personality measures as predictors of work and non-work criteria. *Personality and Individual Differences, 49*, 95–101.

Brannick, M. T. (2008). Back to basics of test construction and scoring. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 131–133.

Brunell, A. B., Gentry, W. A., Campbell, W. K., Hoffman, B. J., Kuhnert, K. W., & DeMarree, K. G. (2008). Leader emergence: The case of the narcissistic leader. *Personality and Social Psychology Bulletin, 34*, 1663–1676.

Callinan, M., & Robertson, I. T. (2000). Work sample testing. *International Journal of Selection and Assessment, 8*, 248–260.

Chan, D. (1996). Criterion and construct validation of an assessment centre. *Journal of Occupational and Organizational Psychology, 69*, 167–181.

Christiansen, N. D., Burns, G., & Montgomery, G. E. (2005). Reconsidering the use of forced-choice formats for applicant personality assessment. *Human Performance, 18*, 267–307.

Christiansen, N. D., Burns, G., & Rozek, R. F. (2010). Effects of socially desirable responding on hiring judgments. *Journal of Personnel Psychology, 9*, 27–39.

Christiansen, N. D., Honts, C. R., & Speer, A. B. (2011, May). *Assessment of personality through behavioral observations in work simulations*. Paper presented at the 15th Conference of the European Association of Work and Organizational Psychology, Maastricht, The Netherlands.

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.

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.

Craik, K. H., Ware, A. P., Kamp, J., O'Reilly, C., Staw, B., & Zedeck, S. (2002). Explorations of construct validity in a combined managerial and personality assessment programme. *Journal of Occupational and Organizational Psychology, 75*, 171–193.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302.

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.

Fleenor, J. W. (1996). Constructs and developmental assessment centers: Further troubling empirical findings. *Journal of Business and Psychology, 10*, 319–335.

Foti, R. J., & Hauenstein, N. M. (2007). Pattern and variable approaches in leadership emergence and effectiveness. *Journal of Applied Psychology, 92*, 347–355.

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

Funder, D. C., Furr, R. M., & Colvin, C. R. (2000). The Riverside Behavioral Q-sort: A tool for the description of social behavior. *Journal of Personality, 68*, 450–489.

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.

Gaugler, B. B., & Thornton, G. C. (1989). Number of assessment center dimensions as a determinant of assessor accuracy. *Journal of Applied Psychology, 74*, 611–618.

Goffin, R. D., Rothstein, M. G., & Johnston, N. G. (1996). Personality testing and the assessment center: Incremental validity for managerial selection. *Journal of Applied Psychology, 81*, 746–756.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84–96.

Haaland, S., & Christiansen, N. D. (2002). Implications of trait-activation theory for evaluating the construct validity of assessment center ratings. *Personnel Psychology, 55*, 137–163.

Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology, 57*, 639–683.

Heller, D., Ferris, D. L., Brown, D., & Watson, D. (2009). The influence of work personality on job satisfaction: Incremental validity and mediation effects. *Journal of Personality, 77*, 1051–1084.

Hoeft, S., & Schuler, H. (2001). The conceptual basis of assessment centre ratings. *International Journal of Selection and Assessment, 9*, 114–123.

Hoffman, B. J., Melchers, K. G., Blair, C. A., Kleinmann, M., & Ladd, R. T. (2011). Exercises and dimensions are the currency of assessment centers. *Personnel Psychology, 64*, 351–395.

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

Jackson, D. J., Stillman, J. A., & Englert, P. (2010). Task-based assessment centers: Empirical support for a systems model. *International Journal of Selection and Assessment, 18*, 141–154.

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.

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology, 87*, 765–780.

Kleinmann, M., Ingold, P. V., Lievens, F., Jansen, A., Melchers, K. G., & König, C. J. (2011). A different look at why selection procedures work: The role of candidates' ability to identify criteria. *Organizational Psychology Review, 1*, 128–146.

Kolk, N. J., Born, M. P., & Van der Flier, H. (2004). Three method factors explaining the low correlations between assessment center dimension ratings and scores on personality inventories. *European Journal of Personality, 18*, 127–141.

Krajewski, H. T., Goffin, R. D., Rothstein, M. G., & Johnston, N. G. (2007). Is personality related to assessment center performance? That depends on how old you are. *Journal of Business and Psychology, 22*, 21–33.

Lance, C. E. (2008). Where have we been, how did we get there, and where shall we go? *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 140–146.

Lievens, F., Chasteen, C. S., Day, E. D., & Christiansen, N. D. (2006). Large-scale investigation of the role of trait activation theory for understanding assessment center convergent and discriminant validity. *Journal of Applied Psychology, 91*, 247–258.

Lievens, F., De Corte, W., & Schollaert, E. (2008). A closer look at the frame-of-reference effect in personality scale scores and validity. *Journal of Applied Psychology, 93*, 268–279.

Lievens, F., De Fruyt, F., & Van Dam, K. (2001). Assessors' use of personality traits in descriptions of assessment centre candidates: A five-factor model perspective. *Journal of Occupational and Organizational Psychology, 74*, 623–636.

Lievens, F., Tett, R. P., & Schleicher, D. J. (2009). Assessment centers at the crossroads: Toward a reconceptualization of assessment center exercises. *Research in Personnel and Human Resources Management, 28*, 99–152.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications [Special issue]. *Journal of Personality, 60*, 175–215.

McFarland, L. A., Ryan, A. M., & Kriska, S. D. (2003). Impression management use and effectiveness across assessment methods. *Journal of Management, 29*, 541–646.

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

Meriac, J. P., Hoffman, B. J., Woehr, D. J., & Fleisher, M. S. (2008). Further evidence for the validity of assessment center dimensions: A meta-analysis of the incremental criterion-related validity of dimension ratings. *Journal of Applied Psychology, 93*, 1042–1052.

Meyer, R. D., Dalal, R. S., & Bonaccio, S. (2009). A meta-analytic investigation into the moderating effects of situational strength on the conscientiousness–performance relationship. *Journal of Organizational Behavior, 30*, 1077–1102.

Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review, 80*, 252–283.

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.

Monahan, E., Hoffman, B. J., Williams, A., & Lance, C. (2012). *A meta-analysis of the validity of assessment center exercises*. Paper presented at the 27th annual SIOP conference, San Diego, CA. **[AQ3]**

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

Neidig, R. D., & Neidig, P. J. (1984). Multiple assessment center exercises and job relatedness. *Journal of Applied Psychology, 69*, 182–186.

Rupp, D. E., Thornton, G. C., & Gibbons, A. M. (2008). The construct validity of the assessment center method and usefulness of dimensions as focal constructs. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 116–120.

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

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262–274.

Schmit, M. J., & Ryan, A. M. (1993). The Big Five in personnel selection: Factor structure in applicant and non-applicant populations. *Journal of Applied Psychology, 78*, 966–974.

Schmitt, N. (1977). Interrater agreement in dimensionality and combination of assessment center judgments. *Journal of Applied Psychology, 62*, 171–176.

Scholz, G., & Schuler, H. (1993). Das nomologischenetzwerk des assessment centers: Einemetaanalyse [The nomological network of assessment centers]. *ZeitschriftfürArbeits und Organisationspsychologie, 37*, 73–85.

Shore, T. H., Thornton, G. C., & Shore, L. M. (1990). Construct validity of two categories of assessment center dimension ratings. *Personnel Psychology, 43*, 101–116.

Spector, P. E., Schneider, J. R., Vance, C. A., & Hezlett, S. A. (2000). The relation of cognitive ability and personality traits to assessment center performance. *Journal of Applied Social Psychology, 30*, 1474–1491.

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.

Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology, 88*, 500–517.

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

Tett, R. P., & Guterman, H. A. (2000). Situation trait relevance, trait expression, and cross-situational consistency: Testing a principle of trait activation. *Journal of Research in Personality, 34*, 397–423.

Thornton, G. C., Tziner, A., Dahan, M., Clevenger, J. P., & Meir, E. (1997). Construct validity of assessment center judgments: Analyses of the behavioral reporting method. *Journal of Social Behavior and Personality, 12*, 109–128.

Van Iddekinge, C. H., Raymark, P. H., & Roth, P. L. (2005). Assessing personality with a structured employment interview: Construct related validity and susceptibility to response inflation. *Journal of Applied Psychology, 90*, 536–552.

Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2005). Is there a general factor in ratings of job performance? A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology, 90*, 108–131.

Wernimont, P. F., & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology, 52*, 372–376.

Woehr, D. J., & Arthur, W., Jr. (2003). The construct-related validity of assessment center ratings: A review and meta-analysis of the role of methodological factors. *Journal of Management, 29*, 231–258.