Texila International Journal of Management
ISSN: 2520-310X
DOI:10.21522/TIJMG.2015.11.02.Art041

A Systematic Review of Assessment Center Validity: Strategic, Technological, and Methodological Advances in Managerial Recruitment

Usha Rudradat
Department of Management, Texila American University, Guyana

Abstract

This study provides the first large-scale examination of the nomological net and incremental validity of assessment center (AC) dimensions identified by. Using two large managerial samples (N = 4,985), investigate the relationships between seven primary AC dimensions and key individual difference variables—cognitive ability and the Big Five personality traits. Findings indicate that AC dimension scores offer meaningful incremental validity beyond traditional psychometric tests. While overall AC ratings showed modest validity (r = .36, mechanically combined composites of AC dimensions produced substantially higher operational validities (unit-weighted r = .44–.50; optimally weighted r = .45–.52), especially when used in conjunction with cognitive and personality assessments. Notably, dimensions such as "problem solving" and "influencing others" showed the strongest associations with cognitive ability and extraversion, respectively, and provided significant incremental gains in predicting job performance (ΔR² = .09–.12). Moreover, single AC dimensions like "problem solving" yielded comparable predictive utility to full composites when paired with personality or cognitive tests. The results underscore the added value of construct-based dimension scoring over holistic AC ratings, which are often diluted by subjective assessor judgments. This research contributes critical evidence supporting the construct validity and applied utility of AC dimension scoring in managerial selection contexts. It also highlights the potential for slimmed-down ACs targeting high-validity dimensions to improve efficiency without sacrificing predictive accuracy.

Keywords: *Assessment Centers, Cognitive Ability, Incremental Validity, Personality Traits.*

Introduction

Assessment centers (ACs) remain a popular tool for the evaluation of job applicants and employees, especially for managerial jobs. As an assessment method, ACs can be designed to measure a multitude of individual differences characteristics (e.g., interpersonal skills, communication skills, personality, cognitive ability). A variety of tools or exercises can be employed to measure these characteristics as part of an AC (e.g., simulations, interviews, in-baskets). Because of the complexities involved in assessing different personal characteristics with a variety of measurement techniques, ACs remain a fruitful field for applied research.

The criterion-related and construct validity of AC ratings has been investigated in many primary studies as well as several meta-analyses. By now there is little doubt that ACs are useful tools for performance prediction. Ref. [12] reported a meta-analytic operational validity of r = .36 for overall AC ratings predicting job performance (corrected for criterion unreliability and range restriction). Two meta-analytic updates of the literature published since 1987 [18, 19]. Report operational validity estimates of r = .28 (corrected for unreliability in the criteria only). These estimates of operational validity for *overall* AC ratings are commonly referred to when ACs are compared with other methods and predictors to estimate their utility in personnel staffing [36].

The issue of AC validity, however, is somewhat

Received: 03.07.2025 ***Accepted: 09.08.2025*** ***Published on: 29.08.2025***
Corresponding Author: udr8321@yahoo.com

more complex. Rather than focusing on validity estimates for overall scores obtained using the AC *method*, researchers and practitioners may benefit from examining the validity of the *constructs* underlying such scores. A construct-based approach provides the opportunity to evaluate the true utility of ACs in applied settings by estimating their incremental validity over other commonly used predictors. As this study will show, a more differentiated validation approach based on AC dimensions, rather than overall AC ratings, will yield more promising estimates of operational and especially incremental validity by considering different predictor combinations.

Recently, [2] took the first step in this direction by conducting a meta-analytic investigation of the primary dimensions underlying AC ratings. Arthur and colleagues argued that estimates of operational validity of overall AC ratings cannot be meaningfully compared with validities of predictors such as personality traits, for example, as this essentially involves comparing an aggregate of different constructs (the overall score across several AC exercises and dimensions) to single constructs (e.g., conscientiousness). Furthermore, comparisons of this sort are comparisons of a *method* (the AC) with *constructs* or *traits* (e.g., cognitive ability, personality), rendering investigations of relative predictive value even less fruitful [3].

To remedy this problem, Arthur and colleagues established a taxonomy of seven primary AC dimensions from a list of 168 lower-order constructs typically measured in ACs [2]. The authors provided meta-analytic estimates of the interrelationships among the primary AC dimensions and showed that they are reasonably related to one another (sample size weighted mean r =.52). More importantly, they also investigated the criterion-related validity for six of these AC dimensions in predicting job performance. Their results demonstrate that the higher-order dimensions commonly assessed across various AC exercises are of different value when predicting overall job performance.

However, to determine the usefulness of AC scores in predicting valued behaviors and outcomes, we must not only assess their criterion- related validity but also their potential overlap with more readily available (and cheaper) assessment tools such as psychometric tests of personality and cognitive ability. Such investigations will result in estimates of incremental validity that AC scores can offer over other tools in the prediction of performance and other valued outcomes. Data that speaks about this matter is available for overall AC ratings. Ref. [6] meta-analytically examined the nomological net of overall AC ratings that represent aggregates of several constructs. Collins and colleagues' meta-analysis pro- vides information on the overlap of overall AC ratings with tests of cognitive ability ($\bar{r}$ = .43) and four of the Big 5 dimensions of personality (agreeableness, extra- version, emotional stability, and openness; $\bar{r}$ = .12, .36, .26, and .18, respectively). Unfortunately, the quantitative summary by Collins and colleagues does not speak to the overlap of construct-based AC *dimensions* with these individual differences' variables. Similarly, an earlier meta-analytic review published in German [38]. Only provided estimates for the relationships of cognitive ability and the Big 5 with overall AC ratings. Equivalent investigations for the construct-based primary dimensions identified by [2] are lacking. Ref. [2] In Table 2 outlining definitions of the seven 'meta-dimensions,' as well as an examination of the lower-order dimensions that contributed to each of these meta-dimensions, provide an indication of which construct-based AC dimensions can be expected to overlap at least partially with individual differences traits. The most obvious associations can be drawn between AC scores on 'problem solving' and cognitive ability. According to Arthur and colleagues, problem solving is the ability to gather information, to effectively analyze data, and to generate viable ideas and solutions to problems. These skills and abilities are at the heart of general mental ability as operationalized by standardized tests [5] and we would thus expect a moderate to strong relationship between AC problem- solving dimension and

cognitive ability test scores. Based on the definition of the six remaining AC dimensions we would not expect much overlap with cognitive ability tests, and thus postulate potential for these AC scores to increase validity of ability measures. However, we stress that cumulative or large-scale evidence regarding these relationships is not available in the published literature, and thus their incremental validity is unknown.

We would also expect that many of the non-cognitive AC dimensions display moderate to strong associations with personality traits as operationalized by standardized tests. The AC dimensions 'influencing others' (conceptually related to leadership skills) and 'drive' (which captures aspects of ambition) should both relate to extraversion, while the latter should also carry some conscientiousness variance, incorporating both achievement and persistence [2]. Other AC dimensions that are easily linked to personality traits on a conceptual level are 'organizing and planning' (which bears conceptual similarity to facets of conscientiousness, especially order) and 'tolerance for stress' (which falls squarely into the domain of emotional stability). Yet again, large-scale evidence for the overlap between these primary AC dimensions and personality traits (whether measured by standardized tests or other methods such as interviews or observer ratings) is lacking.

Based on data presented by [36]. We know that the potential for incremental validity of *overall* AC scores over tests of general mental ability is small (.02). Similarly, the meta-analytic estimates provided by [6] can be used to determine that the incremental validity of *overall* AC ratings over psychometric tests of personality and cognitive ability is negligible. However, the lack of correlational evidence for AC *dimensions* hinders parallel investigations for construct-based AC scores. If correlations between the primary AC dimensions identified by [2] and personality and cognitive ability tests were available, the incremental validity of optimally as well as unit-weighted AC score composites over such tests could be established by multiple regression procedures using the meta-analytic operational validity estimates for all predictors. Additionally, knowledge of these relationships would allow organizations to supplement existing predictor batteries with only a few AC dimensions in a targeted manner, to maximize overall validity efficiently.

The Present Study

The goal of this study is to further AC research by examining the nomological net of the seven primary AC dimensions identified by [2]. First, we provide the first large-scale investigation of the overlap between AC dimensions with individual differences traits (cognitive ability and personality) in two independent samples (total N = 4985). The sample sizes available in our primary data exceed those of prior investigations of overall AC ratings, even of meta-analytic ones [6, 38].

Second, we integrate the findings from our primary data with meta-analytic predictive validity estimates for ACs, personality, cognitive ability, as well as meta-analytic data on the interrelationships of these predictors. The resulting matrix of meta-analytic and large sample primary data provides the best estimates of the relationships between AC dimensions, overall AC ratings, the Big 5 personality dimensions, cognitive ability, and job performance. Subsequently, this matrix of inter-correlations is used to estimate the incremental validity of various combinations of AC scores (overall AC ratings, single dimensions, as well as unit- and optimally weighted dimension composites) over tests of personality and cognitive ability (and vice versa). Thus, the present research addresses four previously unanswered questions:

1. What is the relationship between the seven meaningful primary AC dimensions identified by [2] and the Big 5 personality factors and cognitive ability?
2. What is the incremental validity that can be expected from AC ratings when these primary dimensions are combined (using both

optimal and unit-weights) over tests of cognitive ability and personality?

3. Can useful levels of incremental validity be obtained by adding only one AC dimension to tests of personality and cognitive ability?
4. What is the incremental validity that tests of personality and cognitive ability in turn add over AC dimension composites?

Method

The analytic approach of this study was twofold: first, we strove to provide the best estimate of incremental validity that ACs offer over tests of personality and cognitive ability, based on *meta-analytic estimates* of operational validities and predictor relationships where possible. In this investigation, only the previously unknown inter-correlations between AC dimensions and tests of personality and cognitive ability were estimated from primary data. The second goal was to provide a parallel investigation using *only primary data* on predictor interrelationships, to provide a check on the theoretically derived results as they would occur in a real-world assessment setting.

Meta-analytic Data

To estimate the incremental validity of one predictor over another in predicting a given criterion, the criterion-related validity of each predictor as well as the interrelationship between the predictors must be known. Thus, to inform computations of criterion-related and incremental validity of ACs, we strove to obtain the best validity estimates for AC ratings, the Big 5 personality factors, and cognitive ability, as well as the best estimates of all predictor interrelationships. For this purpose, we consulted the most recent and comprehensive meta-analytic summaries in each research domain.

Next, we obtained sample size weighted mean correlations between all predictors (not corrected for unreliability or range restriction) from these meta- analyses. Where only corrected estimates were available, we obtained the respective artifact distributions and attenuated true validity estimates appropriately.[1] We also obtained estimates of each predictor's criterion-related validity for predicting job performance. For these estimates, operational validities (corrected for unreliability in the criterion and range restriction where appropriate) were obtained where possible. To most accurately estimate the operational validity of a predictor combination in applied settings, it is necessary to use uncorrected estimates of predictor intercorrelations but estimates of criterion-related validity that are corrected for range restriction and unreliability in the criterion measures [36]. The operational validity of predictor combinations with more than two predictors can then be calculated using multiple regression analyses on the full correlation matrix.

In many cases multiple meta-analytic estimates were available. In selecting those that would contribute to our analyses, we took care to match predictors and criterion scales to the current investigation. Since ACs are mostly employed for medium-to-high-person jobs (e.g., to assess managerial applicants), we chose estimates that most closely matched the predictor and criterion relationships of interest. For example, when meta-analytic estimates were reported by sample type, we chose those values that were obtained in managerial samples. Similarly, when values for different criteria were reported, we chose those estimating the predictive validity for managerial performance or in high complexity jobs [20]. for managerial samples, rather than other meta-analytic values available across jobs; or data on high-complexity jobs from [22]. Rather than results across job complexity levels. A detailed description of all sources, as well as the corrections that were applied and the respective artifact distributions used is provided by [6].

Primary Data

In addition to meta-analytic estimates, primary data was employed for two purposes. First, as described earlier, the relationships between [2]. Primary AC dimensions and tests of personality and cognitive ability have not been investigated to date. This study constitutes the first

such investigation by providing estimates obtained from two large primary samples for AC dimensions/personality/cognitive ability inter-correlations. Additionally, data on the interrelationships between personality variables and cognitive ability test scores as well as among AC dimensions were obtained from our two samples. These were used to conduct parallel investigations to those informed by meta-analytic values, to provide a comparison using data obtained from an operational AC.

Data were gathered from AC evaluations of mid-level managers and top-executives conducted for employment purposes[2] over the course of 4 years (2000– 2004). Two different ACs were used to evaluate managers depending on their current managerial level or that of the job they were being considered for. The primary dimensions for which scores were extracted from the two ACs were identical for the two samples and corresponded to those identified by [2].

Sample 1: Mid-level managers

Managers that underwent the mid-level managerial AC (N = 3062) mostly came from first-line or mid-level management and were of an average age of 41.3 years (SD = 7.0) and predominantly male (75.6%). Data on ethnic group membership were available for over 75% of these individuals, 90.2% of whom indicated White, 3.8% Black, 3.5% Hispanic, 1.8% Asian, .02% American Indian, and .5% 'Other' as their ethnic background. Individuals in this sample were well educated, with 48.1% having obtained a Bachelor's degree, and 36.5% a Master's or doctorate/professional degree. Assesses came from over 30 diverse industries (including food processing, electronics manufacturing, wholesale trade, banking and finance, health care, agriculture, construction, government, and transportation), and mostly from large organizations (over 80% of organizations with more than 1000 members, median = 11,000). Man- agers' work experience at the time of assessment was 19.6 years on average (SD = 7.6), in which individuals on average worked for over three employers and had managerial responsibility for a duration of 11.8 years (SD = 7.5). The number of employees managed by the individuals in this sample was as high as 7500, with a median of 18 (median number of direct reports = 6). Individuals received an average annual salary of over US$147,000.

Sample 2: Top-Level Managers

Managers that underwent the top-level managerial AC (N = 1923) were primarily from mid-level to top-executive management, had an average age of 44.3 years (SD = 6.4), and were mostly male (80.4%). Ethnicity data were available for over 78% of these individuals, 92.6% of whom indicated White, 3.0% Black, 1.9% Hispanic, 1.7% Asian, .03% American Indian, and .5% 'Other.' As expected, individuals in this sample on average had an even higher level of education than those in Sample 1, with 42.3% having obtained a Bachelor's degree, and 48.7% a Master's or doctorate/ professional degree. The composition of the sample was like Sample 1 with respect to industries represented, although individuals in this sample tended to come from larger organizations (median = 20,000 members). The average number of years of work experience at the time of assessment was 22.5 (SD = 7.1), with an average of three employers and managerial responsibility of 16.6 years (SD = 7.0). The number of employees managed by the predominantly top-level managers in this sample was as high as 35,000, with a median of 55. Individuals in this sample received an average annual salary of more than US$309,000, with over 10% of managers earning US$500,000 or more.

Measures Used in Primary Data Collection

ACs

Mid-level and top-level managers were evaluated in two different ACs that assessed the same seven higher- order dimensions. The two ACs were designed based on solid scientific principles and current research evidence and conducted by a consulting company with extensive experience in assessment for selection and development

purposes (serving 75 of the Fortune 100 companies). Both ACs used background interviews, an in-basket exercise, and three role-play exercises in the form of a direct report meeting, a strategy presentation, and a task force meeting. Depending on the exercise, individuals were assessed by one or multiple trained and experienced assessors. Assessors evaluated managers' performance on several lower-order competencies within each exercise. Ratings were provided on very detailed behaviorally anchored rating scales. Across exercises in each AC, individual managers were rated on several hundred distinct behaviors that were combined to derive competency scores. Subsequently, AC primary dimension scores were obtained by summing relevant competency scores across AC exercises. The resulting seven primary dimensions were communication, consideration/aware- ness of others, drive, influencing others, organizing and planning, problem solving, and tolerance for stress/ uncertainty [31].

Cognitive Ability

Cognitive ability was assessed using three different psychometric tests: the Watson–Glaser Critical Thinking Appraisal [40]. The Wesman Personnel Classification Test [33, 32] and the Raven's Progressive Matrices [34]. All three inventories are widely used tests of cognitive ability that see heavy use in personnel selection settings and are supported by extensive evidence demonstrating their construct validity as well as reliability. Individuals' scores were transformed to standardized scores based on the normative information available for each measure and were subsequently combined to obtain an overall score as an indicator of general mental ability.

Personality

Personality was assessed using the Global Personality Inventory [10]. A personality inventory that was developed to assess the Big 5 personality factors emotional stability, extraversion, openness to experience, agreeableness, and conscientiousness. Facet scales that assess each of the Big 5 factors include, among others, Emotional Control, Optimism, and Stress Tolerance (for emotional stability), Competitive- ness, Desire for Achievement, and Sociability (for extraversion), Independence, Creativity, and Vision (for openness), Consideration, Empathy, and Trust (for agreeableness), and Attention to Detail, Dutiful- ness, and Responsibility (for conscientiousness). The GPI was developed especially for use in employment settings. Extensive evidence exists on its reliability and construct validity across diverse jobs as well as from countries around the world [31, 37].

Analyses

Three separate analyses were carried out to arrive at different estimates for operational and incremental validity. First, the meta-analytic data on predictor inter-correlations and criterion-related validity were used in combination with the primary data on AC dimension–predictor relationships to obtain the best estimate of operational validity by means of multiple regression as well as composite correlations. These analyses involved computations of operational validities for AC dimensions as a set, AC dimensions when combined with personality and cognitive ability, and tests of personality and cognitive ability combined. The process was repeated for an optimally weighted composite of AC dimensions as well as for a unit-weighted composite.

Unit-weighted estimates were obtained by weighing all AC dimensions equally using the formula for computing composite correlation [13, 29]. While optimally weighted estimates were obtained using regression weights. In estimating the joint validity of unit-weighted AC dimension composites and other predictors, Big 5 and cognitive ability were optimally weighted based on their meta-analytic regression weights. This procedure in turn required us to first compute composite correlations among AC dimensions with each individual differences' predictor, as well as multiple correlations between predictors and the performance criteria, and then subject the resulting correlation

matrix to multiple regression analyses. Results from these analyses were then used to compute incremental validity of each predictor over all possible predictor combinations.

Second, the same analyses were carried out for each single AC dimension to evaluate whether useful levels of incremental validity could be obtained using only one, rather than all, AC dimensions. Third, parallel analyses were carried out using primary data on predictor inter-correlations and meta-analytic data for operational validity in the sample of 3062 mid-level managers (criterion information was not available in the primary data). Finally, the same analyses were repeated using the data obtained from the sample of 1923 top- level managers.

Our approach is unique as it yields results that have so far not been available in the literature: even though the operational validity of primary AC dimensions has previously been investigated, their relationships with other predictors were unknown, and incremental validities for AC *dimension* score composites were thus unavailable.

Results

AC Dimension Overlap with Cognitive Ability and Personality

The observed inter-correlations among AC dimensions, Big 5 personality factors, and cognitive ability are presented in Table 1. Again, the data presented here are the first estimates of this kind, relating to the over-arching AC dimensions identified by [2] to individual differences in personality and cognitive ability. Correlations from the mid-level and top-manager samples are presented separately as well as combined (sample size weighted). For comparative purposes, the correlations between cognitive ability and the Big 5 with overall AC ratings from [6, 38] are also given in Table 1.

The correlational pattern was remarkably similar across the two samples of managers, and hence only the results for the combined sample are discussed here. The pattern of relationships between AC dimensions, cognitive ability, and the Big 5 was very compatible with the definitions of the seven AC dimensions offered by [2]. As expected, the AC dimensions 'problem solving' dis- played the strongest relationship with cognitive ability (r =.32) and the personality trait openness to experience (r =.18). This pattern is intuitively appealing, as problem solving in ACs has been postulated to describe the ability to analyze and reason with information, as well as the ability to generate ideas and imaginative solutions [2]. Also as expected, problem solving was the only one of the seven AC dimensions that displayed a notable relationship to cognitive ability. Other AC dimensions displayed sizable relationships with personality traits, however, and mostly confirmed our expectations.

'Organizing and planning' related most strongly to the personality domain of conscientiousness (r =.24). This finding is not surprising – orderliness has been postulated as a facet of conscientiousness in many personality taxonomies and has also been empirically established as such [35]. The AC dimension 'influencing others,' which we expected to relate mostly to extraversion, was found to relate to several of the Big 5 personality domains. It indeed displayed a sizable relationship with extraversion (observed r =.27) but also related to agreeableness (r =.27) and emotional stability (r =.24). Influencing others describes an individual's persuasiveness and urgency. Indeed, most of the lower-order competencies that [2]. Categorized into this primary dimension relates to leadership skills. This was also the case in the present two samples, where many of the lower-order competencies that were included in this dimension related to leading and developing others. Thus, the pattern of relationships with the Big 5 factor of extraversion, but also emotional stability and agreeableness, is in line with the broad leadership literature. These three personality traits relate to leadership effectiveness in general [23] as well as transformational leadership [4].

'Consideration/awareness of others mostly reflects social skills and teamwork-related competencies. Appropriately, these AC

dimensions related most strongly to the Big 5 domain of agreeableness (r =.27). However, it also reflected individuals' emotional stability to some extent (r =.20). 'Communication' was the only AC dimension that displayed virtually no notable relationship with any of the individual differences' scales included in this study. The strongest relationship was found to exist between this AC dimension and agreeableness; however, the relationship was quite weak (r =.08). Unlike communication, the dimension of 'drive' was found to relate to all personality scales.

The multitude of relationships as well as their strengths were greater than expected; based on the construct definition this dimension was conceptually only associated with high extraversion and high conscientiousness. While drive related most strongly to extraversion (r =.48), it also correlated notably with the remaining Big 5 traits (r =.31, .32, .26, and .31 with emotional stability, openness, agreeableness, and conscientiousness, respectively). *Post hoc*, we can make sense of these relationships with the Big 5 by examining the lower- order competencies that define drive as a primary AC dimension. The taxonomy was established by [2] lists a number of diverse characteristics- aggressiveness, perseverance, and initiative as belonging to this category. Thus, it is not surprising that drive would also be related to several of the Big 5 personality traits. Lastly, confirming our expectations, tolerance for stress' exhibited its strongest relationship with emotional stability (r =.50) but was also related to extraversion, albeit to a lesser extent (r =.33).

These results offer a unique contribution to literature and now provide the basis for a better understanding of the construct-underpinnings of AC dimensions. It is important to note that for the purpose of this study, observed (uncorrected) estimates of predictor inter-correlations were of interest. If one were to correct the values presented in Table 1 for attenuation due to unreliability in both types of measures using mean reliability from meta-analyses and reliability generalization studies [8, 21, 39]. One would obtain estimates of construct-level relationships that would be higher than those observed on the measure-level. That is, the relationships between AC dimensions and Big 5 personality traits would on average be 35% larger, while relationships between AC dimensions and cognitive ability would be on average 26% larger than observed values. However, given the applied focus of this study to understand the relative usefulness of AC dimension vis- a`-vis individual differences measures used in personnel selection, our focus on observed predictor interrelationships is appropriate.

Table 1. Correlations Primary Assessment Center Dimensions with Cognitive Ability and Personality

Dimension	*g*		ES		E		O		A		C	
	r	*N*	*r*	*N*	*r*	*N*	*r*	*N*	*r*	*N*	*r*	*N*
Sample 1: Mid-level managers												
Problem solving	.34	2995	.09	2924	.07	2924	.18	2924	.06	2,924	—.07	2924
Organizing and planning	.02	2995	.16	2924	.17	2924	.16	2924	.15	2924	.27	2924
Influencing others	.02	2995	.22	2924	.25	2924	.17	2924	.26	2924	.09	2924
Consideration of others	—.06	2994	.20	2923	.16	2923	.07	2923	.25	2923	.08	2923
Communication	.03	2995	.06	2924	.02	2924	.01	2924	.10	2924	—.05	2924
Drive	—.09	2995	.34	2924	.50	2924	.34	2924	.29	2924	.35	2924
Tolerance of stress	.02	2994	.49	2924	.22	2924	.10	2924	.22	2924	.11	2924
Unit-weighted composite	.07	2995	.26	2924	.28	2924	.23	2924	.26	2924	.16	2924
Sample 2: Top-level managers												

Problem solving	.28	1861	.08	1854	.17	1854	.19	1854	.04	1854	—.03	1854
Organizing and planning	.10	1861	.11	1854	.17	1854	.11	1854	.07	1854	.20	1854
Influencing others	.04	1860	.26	1853	.29	1853	.21	1853	.28	1853	.11	1853
Consideration of others	.02	1861	.21	1854	.16	1854	.05	1854	.29	1854	.06	1854
Communication	.09	1859	.02	1852	—.02	1852	—.03	1852	.06	1852	—.06	1852
Drive	.03	1861	.25	1854	.44	1854	.29	1854	.20	1854	.25	1854
Tolerance of stress	.03	1861	.52	1854	.50	1854	.29	1854	.27	1854	.25	1854
Unit-weighted composite	.13	1878	.20	1854	.24	1854	.17	1854	.18	1854	.13	1,854
Both samples combined (sample size weighted)												
Problem solving	.32	4856	.09	4778	.11	4,778	.18	4778	.05	4778	—.05	4778
Organizing and planning	.05	4856	.14	4778	.17	4,778	.14	4,778	.12	4778	.24	4778
Influencing others	.03	4855	.24	4777	.27	4777	.19	4777	.27	4777	.10	4777
Consideration of others	—.03	4855	.20	4777	.16	4777	.06	4777	.27	4777	.07	4777
Communication	.05	4854	.04	4776	.00	4776	—.01	4776	.08	4776	—.05	4776
Drive	—.04	4856	.31	4778	.48	4778	.32	4778	.26	4778	.31	4778
Tolerance of stress	.02	4855	.50	4778	.33	4778	.17	4778	.24	4778	.16	4778
Unit-weighted composite	.09	4873	.24	4778	.26	4778	.21	4778	.23	4778	.15	4778
Overall AC ratings[a]	.43[b]	5419	.26[b]	1023	.36[b]	1847	.18[b]	619	.12[b]	830	.14[c]	1107

Note: *The unit-weighted composite is a sum of the first six dimensions only, in order to provide data consistent with the computations of operational validities, which were only available for these six dimensions from [2].*

[a]Meta-analytic values for overall AC ratings are presented for comparison purposes. The respective values were obtained from the most recent meta-analytic summaries. For relationships where estimates were not available in the most recent quantitative review, prior meta-analytic work was consulted.

[b]Sample size weighted mean *r* from [6].

[c]Sample size weighted mean *r* for conscientiousness and achievement motivation (weighted by total *N*) from [38] *g*, general mental ability (overall score on three cognitive ability tests); ES, emotional stability; E, extraversion; O, openness; A, agreeableness; C, conscientiousness.

Inter-correlations Among AC Dimensions

Observed inter-correlations among AC dimensions from our two samples are presented in Table 2. Results show that the seven dimensions were moderately related ($\bar{r}$ = .33 and .34 in the mid-level and top-level manager samples, respectively). These values are noticeably lower than those reported in the quantitative summary of [2]. where the average sample size weighted mean *r* among the seven primary AC dimensions was reported as .52. The lower average correlation between the AC dimensions in the primary data may lead to higher estimates of AC validity once dimension composite validities are estimated. One should note that the sample sizes for many of the AC dimension inter-correlations obtained from the primary data in this study are much larger than

the cumulated sample sizes for the meta-analytic values reported in [2]. For nine out of the 15 correlations, the sample sizes in this study exceed those of Arthur and colleagues; for eight of these nine correlations, the sample sizes were more than twice as large as the meta- analytic *N*s. Although meta-analytic data has advantages over single sample investigations, well-conducted, large-scale primary data can add significant value, especially in cases like the one at hand, where the meta-analytic investigation does not include much of the unpublished work

Table 2. Inter-correlations among Primary Assessment Center Dimensions in Samples 1 and 2

Dimension	1	2	3	4	5	6	7
1. Problem solving		.50	.41	.20	.25	.40	.28
n		1923	1922	1923	1921	1923	1922
2. Organizing and planning	.42		.42	.17	.21	.40	.27
n	3062		1922	1923	1921	1923	1922
3. Influencing others	.46	.45		.53	.44	.44	.41
n	3062	3062		1922	1920	1922	1921
4. Consideration/awareness of others	.27	.26	.52		.63	.15	.27
n	3061	3061	3061		1921	1923	1922
5. Communication	.33	.27	.45	.60		.12	.17
n	3062	3062	3062	3061		1921	1920
6. Drive	.25	.40	.38	.27	.14		.46
n	3062	3062	3062	3061	3062		1922
7. Tolerance of stress/uncertainty	.23	.20	.33	.31	.20	.25	
n	3061	3061	3061	3060	3061	3061	

***Notes*:** Values below diagonal obtained from Sample 1 (mid-level manager assessment center); values above diagonal from Sample 2 (top-level manager assessment center).

Validity of AC Dimensions

Operational Validity

As described in the analyses section, meta-analytic validities for the six AC dimensions' communication, consideration/awareness of others, drive, influencing others, organizing and planning, and problem solving were obtained from [2]. These estimates, together with observed inter-correlations among AC dimension scores (both meta-analytic and from primary data), were used to obtain multiple correlations between AC dimensions and overall job performance. The resulting multiple correlations estimate the validity that can be expected from composites of these six AC dimensions. To this end, we computed validities for both unit-weighted and optimally weighted compo- sites of AC dimensions. These validity estimates are presented in Table 3.

The operational validity estimates computed for AC dimension composites based on meta-analytic inter-correlations among the primary dimensions were nearly identical (.44 for unit-weight and .45 for optimally weighted composites). The same was the case for validity estimates computed based on AC dimension inter-correlations obtained from primary data. However, as expected, these validities were slightly higher due to the lower inter-correlations of AC dimensions in the primary data. For the mid-level manager sample, the values were .49 and .52, and for the top-level manager sample .50 and .51 (unit- and optimally weighted, respectively). In Table 3, the validity of overall AC ratings from a prior meta-analysis is presented for comparative purposes (r = .36) [12]. Recall that AC dimension composites that we computed were derived entirely mechanically. Thus, we observe that both unit- and optimally weighted, mechanically combined AC dimension

composites yield validity estimates well above those for overall AC ratings, which are often obtained, at least in part, using subjective information combination on the part of assessors.

Incremental Validity of AC Dimensions Over Personality and Cognitive Ability

Using the information obtained from this study together with meta-analytically derived predictor inter- correlations, we formed predictor composites of AC dimensions (unit- and optimally weighted), personality test scores, and cognitive ability. The resulting validity estimates were substantial. Adding the primary AC *dimensions* (unit-weighted) as a set to the Big 5 factors of personality yields criterion-related validity estimates of .47 for the prediction of job performance (based on meta-analytic predictor inter-correlations), while adding them to measures of cognitive ability increases validity estimates to .68. Combining all three types of predictors available yields an estimate of operational validity of .71. Optimally weighted AC dimension composites performed even slightly better. Similar results were obtained when substituting meta-analytic predictor inter-correlations with those obtained in the primary samples. The detailed computations of operational validities of predictor combinations can be found in [1, 2, 6, 9, 12, 20, 22, 36, 38]. for the meta-analytic and primary samples, respectively. Next, we focus our discussion on incremental validity results.

Incremental validity estimates of AC dimensions over personality and cognitive ability were computed based on the multiple correlations reported above; these results are also reported in Table 3. While the incremental validity that can be expected from *overall* AC ratings over personality measures and tests of cognitive ability has already been reported elsewhere (incremental validity over personality being substantial, [14]. Whereas the incremental validity over cognitive ability is negligible, [17, 37]. The estimates of relative incremental value offered by AC *dimensions* and different ways of combining them are unique to the present study. Results show that mechanically (unit- or optimally) weighted composites of AC dimensions provide incremental validity over tests of cognitive ability and personality. Incremental validity estimates from meta-analytic data were .17 and .27 for unit- and optimally weighted AC dimension composites over tests of personality. The incremental validity of AC dimensions over tests of cognitive ability was estimated at .12/.15 (unit-/optimally weighted). Finally, the incremental validity over *both* types of individual differences predictors was found to be .09/.12. This contrasts with the findings for overall AC ratings, which provide no incremental validity over tests of personality and cognitive ability. That is, regarding operational and incremental validity, both unit- and optimally weighted AC dimension composites fare much better than overall AC ratings (which are often derived by subjectively combining information across AC dimensions or making holistic judgments).

It is encouraging that the estimates obtained from primary data were very similar. The estimates for operational validities of AC dimension composites were only slightly higher due to the lower average inter-correlations among dimensions, and the overall conclusions regarding incremental validity remain unchanged. The incremental validity estimates from our primary data were .11/.13 and .10/.12 for both types of AC composites in the mid-level and top-manager samples, respectively. These values are perfectly in line with the estimate of .09/.12 obtained from the meta-analytically based data.

Table 3. Incremental Validity of Assessment Center Dimensions Over Psychometric Tests for Predicting Job Performance

	Overall AC ratings (incl. subjective combination)	**AC dimensions (unit-weighted)**	**AC dimensions (optimally weighted)**
Meta-analytic estimates			
Operational validity	.36	.44	.45
Incremental validity over.			
Big 5	.12	.17	.27
Cognitive ability	.02	.12	.15
Cognitive ability + Big 5	.00	.09	.12
***Mid-level manager sample* (*N* = 3062)**			
Operational validity		.49	.52
Incremental validity over.			
Big 5		.26	.31
Cognitive ability		.16	.19
Cognitive ability + Big 5		.11	.13
***Top-level manager sample* (*N* = 1923)**			
Operational validity		.50	.51
Incremental validity over			
Big 5		.27	.30
Cognitive ability		.14	.15
Cognitive ability + Big 5		.10	.12

Note: Computations were based on operational validity estimates and observed predictor inter-correlations from sources listed in [6].

Incremental Validity of Single AC Dimensions

The incremental validity that can be expected from six of the seven primary AC dimensions is high enough to consider ACs useful tools for evaluating managerial candidates. However, the question remains whether similar incremental validity for performance prediction can be achieved using only one of the seven AC dimensions. This issue is particularly interesting from an applied perspective. Organizations that are already employing a specific combination of predictor measures may be interested in adding only AC dimensions that supplement their current tools in the most optimal yet efficient manner. The results of our analysis that speak to this issue are presented in Table 4.

Naturally, which specific dimension offers the highest incremental value depends on the combination of predictors already being employed for a given criterion. If one were looking to supplement a test of the Big 5 domains of personality in performance prediction by assessing only a single AC dimension, the best choice regarding incremental validity would be *problem solving*. This comes as no surprise when considering that this AC dimension is by far the most cognitively loaded of the seven primary dimensions (sample size weighted mean observed *r* in the present two samples = .32). Because of its high operational validity (= .39; and universally small overlap with the Big 5 domains of personality (mean *r* in the present two samples = .08) [2]. this AC domain increments validity by .19 correlational points when added to the Big 5.

Conversely, when looking to increment the

validity of tests of cognitive ability, the best choice is the AC dimension *influencing others*. This leadership-related dimension displayed low overlap with cognitive ability scores in our two samples (mean r =.03). Because of its high operational validity (= .38), this dimension would add .11 in incremental validity when added to tests of cognitive ability. This dimension also remains the best choice when tests of cognitive ability and personality are combined with only one AC dimension; the incremental validity estimate is .09 as obtained from meta-analytic data and primary samples. This certainly compares quite favorably with the incremental validity that would be obtained if all AC dimensions were added to the equation; that meta-analytic estimate was only slightly higher (.09/.12; see Table 3).

Table 4. Incremental Validity of each Assessment Center Dimension over Psychometric Tests for Predicting Job Performance AC Dimensions

	Problem solving	**Organizing and planning**	**Influencing others**	**Consideration/ awareness**	**Communication**	**Drive**
Operational validity	.39	.37	.38	.25	.33	.31
Meta-analytic estimates **Incremental validity over**						
Big 5	.19	.12	.14	.07	.16	.05
Cognitive ability	.04	.10	.11	.06	.08	.09
Cognitive ability + Big 5	.04	.07	.09	.04	.07	.07
Mid-level manager sample **(*N* = 3062)** **Incremental validity over**						
Big 5	.24	.15	.19	.09	.17	.09
Cognitive ability	.04	.11	.11	.07	.08	.11
Cognitive ability + Big 5	.04	.06	.08	.05	.08	.05
Top-level manager sample **(*N* = 1923)** **Incremental validity over**						
Big 5	.22	.15	.19	.10	.17	.10
Cognitive ability	.05	.08	.10	.05	.07	.07
Cognitive ability + Big 5	.05	.05	.09	.04	.07	.03

Note: Computations were based on operational validity estimates and observed predictor inter-correlations from sources listed in [6].

Incremental Validity of Personality and Cognitive Ability Over AC Scores

While the focus of the present investigation was on the incremental validity of ACs over tests of personality and cognitive ability, we provide a parallel investigation for the value that such tests offer over overall AC ratings and dimension composites. These results are presented in Table 5; again, details on the computations can be found in [1, 2, 6, 9, 12, 20, 22, 36, 38]. The operational validity of the Big 5 as a set, cognitive ability, and the Big 5 and cognitive ability combined were estimated at .29, .56, and .62 based on the available meta-analytic evidence. Incremental validities were computed based on the correlations with AC dimensions obtained from this study and the meta- analytic correlations with overall AC ratings from [6, 38]. Results show that cognitive ability and the Big 5 as a set add substantial value over all types of AC ratings. The

combined incremental validity of these two predictors over unit- and optimally weighted AC dimension composites was .28, respectively. Similar estimates were obtained from the data based on the two primary samples; the respective estimates were .28 (mid-level manager sample) and .26 (top-level manager sample).

Table 5. Incremental Validity of Psychometric Tests Over Assessment Center Dimensions for Predicting Job Performance

	Big 5	**Cognitive ability**	**Big 5 and cognitive ability**
Meta-analytic estimates			
Operational validity	.29	.56	.62
Incremental validity over			
Overall AC ratings	.05	.22	.26
AC dimensions (unit-weighted)	.03	.25	.27
AC dimensions (optimally weighted)	.11	.25	.29
***Mid-level manager sample* (*N* = 3062)**			
Operational validity	.26	.56	.67
Incremental validity over			
AC dimensions (unit-weighted)	.03	.23	.28
AC dimensions (optimally weighted)	.06	.23	.28
***Top-level manager sample* (*N* = 1923)**			
Operational validity	.26	.56	.66
Incremental validity over			
AC dimensions (unit-weighted)	.03	.20	.26
AC dimensions (optimally weighted)	.05	.20	.26

Note: Computations were based on operational validity estimates and observed predictor inter-correlations from sources listed in [6].

Discussion

This study provided the first investigation of the nomological net of the seven primary AC dimensions established by [2] by assessing their overlap with tests of cognitive ability and personality. The data on these relationships were used in conjunction with prior meta-analytic data on the validity of overall AC ratings, AC dimensions, personality, and cognitive ability, to evaluate the relative incremental value that AC dimension scores add to tests of personality and cognitive ability.

ACs can increase the applied utility of managerial staffing systems by offering information about individuals that supplements data already provided by standardized psychometric tests. Previously, investigations of incremental validity were only available for overall AC ratings, resulting in low to negligible estimates. As this study has shown, a composite of dimension scores offers useful levels of incremental validity when combined with scores of cognitive ability and personality

tests (D*R* = .12 when optimally weighted). However, incrementing utility in such a way requires making use of information on AC dimensions by mechanical combination of primary dimensions rather than using overall AC ratings. The incremental validity of overall AC ratings over tests of personality plus cognitive ability was found to be .00.

A major factor that hampers the criterion-related validity of overall AC ratings in operational settings is the fact that most often scores are combined in a non- optimal fashion across raters, exercises, and dimensions. Such non-

optimal approaches include assigning overall scores based on clinical data combination or discussion among raters [15, 16, 28]. Previous research has shown that mechanically combining AC ratings is superior to judgmentally combining assessor ratings [11]. This research shows that this is also the case when it comes to incremental validity. The mode of data combination does not seem to matter much in relation to validity, if it is mechanical. The operational and incremental validity estimates established in this study were not appreciably different for unit- and optimally weighted AC dimension composites. This was the case with our investigations based on meta- analytic as well as primary data.

From an applied perspective, this study yields an interesting conclusion about whether all AC dimensions need to be assessed to achieve good incremental validity. Given the fact that ACs are costly tools and time consuming to administer [6]. The question of whether useful levels of incremental validity can be achieved with fewer dimensions is worth some consideration. To this end, we conducted parallel analyses estimating the incremental validity of each AC dimension over the Big 5, cognitive ability, and both predictor types combined. The present set of results (from both meta-analytic and primary data) yield a straightforward answer that may be of high utility in applied settings: useful levels of incremental validity can be obtained even when using only one AC dimension. Practitioners looking to design and implement slimmed-down ACs concentrating on only a few dimensions should concentrate on assessing *problem solving* and *influencing others*. These dimensions offer the largest incremental value when added to tests of cognitive ability, personality, or both. The challenge in designing such ACs, however, lies in designing exercises targeted at only a few dimensions, as well as in training assessors to capture and rate only assessed behavior relevant to the specific dimension(s) identified to add incremental validity. It is conceivable that a slimmed- down AC designed to measure only a single AC dimension instead results in assessors providing an overall evaluation, or at least an AC dimension rating contaminated by an overall impression of the candidate. Such scores would likely function similarly to overall AC ratings that are derived using subjective data combination across several dimensions. However, in well-designed and implemented ACs, significant re- sources are typically invested in training assessors and providing them with scientifically sound guidelines and tools for evaluating behaviors observed among candidates. Thus, we are optimistic that such slimmed-down, one- or two-dimension ACs can be developed and implemented for organizations looking to supplement their existing assessment tools. However, because these ACs would put a strong emphasis on maximizing overall validity, they may not be suited to provide developmental feedback. It also remains to be seen whether such slimmed-down ACs really differ from other selection tools that include simulation-type assessments, such as work sample tests, for instance.

Nonetheless, validity estimates for single AC dimensions as well as composites of AC primary dimensions are encouraging. This research has shown that previous estimates of the operational and incremental validity estimates were hampered by the fact that only overall AC ratings were considered, rather than scores on meaningful AC dimensions. We proposed that overall AC scores do not add incremental validity since they are often obtained in a non-optimal fashion (typically using subjective means of data combination). Yet, another major reason is that overall AC scores are often not aligned with other tools that are already being administered. Additional measures increase the overall reliability of a predictor battery and will thus increase overall validity. However, in addition to increasing overall reliability of an assessment battery, construct-based AC dimensions can also add *construct coverage*. For predictive purposes, the more indicators of a given construct we can administer, the better. Each measure will conceptualize the construct domain slightly differently (after all, this study also showed that AC dimensions are not correlated perfectly even

with tests of conceptually related individual differences traits). We now know that the effects of increased construct coverage can go beyond those of simply increasing reliability [7]. Showed that adding a test of conscientiousness to another test of conscientiousness can result in large increases in overall validity, simply because the coverage of the predictor construct increases (albeit the overall validity asymptotes to a certain level after a certain number of scales). They found that this was even the case on facet level, where construct definitions were identical. Similarly, construct-based AC dimensions provide the opportunity to explain variance in job performance with the help of predictor constructs that are not yet well represented or assessed incompletely in each test battery.

As discussed above, we can also add AC dimensions that measure constructs not at all represented in certain predictor batteries. For example, if an organization only uses ability and experience-related constructs to identify suitable candidates, an AC that places a strong emphasis on 'drive' has the potential to add incremental validity. We know from the I/O psychology literature that this is not the case for overall scores [36]. Possibly due to their high correlation with cognitive ability [6]. Thus, there is a need to employ construct-based dimensions that clearly tease predictor constructs apart and show little overlap with exactly those predictors already in use. Again, an emphasis on *constructs* assessed in ACs, rather than on scores for the overall *method*, provides a more fruitful avenue for investigations of incremental validity.

Considering the long-lasting controversy on AC construct validity (for a recent exchange, [25, 8]. and other replies published in *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 1, 1), one might ask the question whether similarly encouraging results of incremental validity would be obtained for individual AC *exercise* scores than those we established here for AC dimensions. While such question has merit, an investigation of AC exercise score validity over individual differences traits would again constitute an investigation of *methods* (AC exercises regardless of dimension assessed) vs *constructs* (e.g., personality or GMA). We are convinced that the potential for incremental validity of AC dimensions as established in our research is based on additional predictor construct coverage, and the opportunity to add individual AC dimensions that assess constructs not yet well assessed by other predictors. It is possible that exercise scores that summarize performance across several predictor constructs function similarly to overall AC ratings. Nonetheless, we see some potential for incremental validity of AC exercise scores when they are combined mechanically such as unit-weighted composites (compared with previous investigations of overall AC ratings that often include subjective methods of data combination). Un- fortunately, investigations of AC exercise validity parallel to the one presented here for AC dimensions are currently impossible – the literature is lacking good validity estimates of AC exercises in predicting job performance. We could only locate a single sample of 359 candidates for first-line supervisor positions Study [2, 26]. That contains correlations between AC exercise factors and a job performance criterion. Even investigations of AC exercise validity for distal performance outcomes such as salary are extremely rare [27]. In this study, we were fortunate enough to be able to rely on meta-analytic evidence on AC dimension criterion-related validity; our field should take strides to accumulate similar data on the usefulness of AC exercises as an additional step toward resolving the AC construct validity question.

Notes

1. Predictor inter-correlations should not be corrected for attenuation due to unreliability, because in estimating operational validities we are interested in the predictive efficiency of these measures as they are used in operational settings, recognizing that scores are subject to measurement error. Estimating operational validity in this way is standard procedure in personnel selection re- search,

and the same procedure is often applied in meta-analyses of criterion-related validities of popular predictors [24, 33, 36].

2. Assessment centers were conducted both for developmental as well as selection purposes. However, the predictor inter-correlations contributing to our analyses were very similar for developmental and selection ACs. Thus, overall findings are virtually identical across assessment purposes. For reasons of brevity, we present only results for the combined analysis.

Acknowledgements

I would like to express my sincere gratitude to Dr. Saranya Williams and Mr. Surendra Boodhoo for their invaluable guidance, insightful feedback, and unwavering support throughout this research. Their combined knowledge base in Strategic Human Resource Management greatly enriched this study. Especially to Mr. Surendra Boodhoo for his constructive suggestions and practical insights.

I am also grateful to the researchers and institutions whose work formed the foundation of this review. Finally, I appreciate the encouragement from colleagues and family, whose support was instrumental in completing this work.

References

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

[2]. 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.

[3]. 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.

[4]. Bono, J. E., & Judge, T. A., 2004, Personality and transformational and transactional leadership: A meta-analysis. *Journal of Applied Psychology*, 89, 901–910.

[5]. Carroll, J. B., 1993, Human cognitive abilities: A survey of factor-analytic studies. *New York: Cambridge University Press.*

[6]. 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.

[7]. Connelly, B. S., & Ones, D. S., 2007, Combining conscientiousness scales: Can't get enough of the trait, baby. *Poster presented at the Society for Industrial and Organizational Psychology Conference*, New York City, NY, April.

[8]. Connelly, B. S., Ones, D. S., Ramesh, A., & Goff, M., 2008, A pragmatic view of dimensions and exercises in assessment center ratings. *Industrial and Organizational Psychology*, 1, 127–130.

[9]. Dilchert, S., & Ones, D. S., 2009, Assessment center dimensions: Individual differences correlates and meta-analytic incremental validity. *International Journal of Selection and Assessment*, 17(3), 254–270, https://doi.org/10.1111/j.1468-2389.2009.00468.x

[10]. ePredix., 2001, Global Personality Inventory Technical Manual. *Minneapolis, MN: ePredix.*

[11]. Feltham, R., 1988, Assessment centre decision making: Judgmental vs. mechanical. *Journal of Occupational Psychology*, 61, 237–241.

[12]. 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.

[13]. Ghiselli, E. E., Campbell, J. P., & Zedeck, S., 1981, Measurement theory for the behavioral sciences. *San Francisco: Freeman.*

[14]. 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.

[15]. Grove, W. M., & Meehl, P. E., 1996, Comparative efficiency of informal and formal prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, 2, 293–323.

[16]. Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C., 2000, Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, 12, 19–30.

[17]. Hardison, C. M., 2005, Construct validity of assessment center overall ratings (Doctoral dissertation). University of Minnesota, *Minneapolis, MN.*

[18]. Hardison, C. M., & Sackett, P. R., 2004, Assessment center criterion-related validity: A meta-analytic update. Paper presented at the Society for Industrial and Organizational Psychology Conference, *Chicago, IL.*

[19]. 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.

[20]. Hough, L. M., Ones, D. S., & Viswesvaran, C., 1998, Personality correlates of managerial performance constructs. Poster presented at the Society for Industrial and Organizational Psychology Conference, *Dallas, TX, April.*

[21]. Hülsheger, U. R., Maier, G. W., & Stumpp, T., 2007, Validity of general mental ability for the prediction of job performance and training success in Germany: A meta-analysis. *International Journal of Selection and Assessment*, 15, 3–18.

[22]. Hunter, J. E., 1980, Validity generalization for 12,000 jobs. *Washington, DC: US Department of Labor.*

[23]. 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.

[24]. Kuncel, N. R., Hezlett, S. A., & Ones, D. S., 2001, A comprehensive meta-analysis of the predictive validity of the GRE: Implications for graduate student selection. *Psychological Bulletin*, 127, 162–181.

[25]. Lance, C. E., 2008, Why assessment centers do not work the way they are supposed to. *Industrial and Organizational Psychology*, 1, 84–97.

[26]. Lance, C. E., Newbolt, W. H., Gatewood, R. D., Foster, M. R., French, N. R., & Smith, D. E., 2000, Assessment center exercise factors represent cross-situational specificity, not method bias. *Human Performance*, 13, 323–353.

[27]. Lievens, F., Dilchert, S., & Ones, D. S., in press, The importance of exercise and dimension factors in assessment centers: Simultaneous examinations of construct- and criterion-related validity. *Human Performance.*

[28]. Meehl, P. E., 1954, Clinical vs. statistical prediction: A theoretical analysis and a review of the evidence. *Minneapolis, MN: University of Minnesota Press.*

[29]. Nunnally, J. C., 1978, Psychometric theory (2nd ed.). *New York: McGraw-Hill*.

[30]. Ones, D. S., 1993, The construct validity of integrity tests (Doctoral dissertation). University of Iowa, *Iowa City, IA.*

[31]. Ones, D. S., Dilchert, S., Viswesvaran, C., & Judge, T. A., 2007, In support of personality assessment in organizational settings. *Personnel Psychology*, 60, 995–1027.

[32]. Ones, D. S., Viswesvaran, C., & Reiss, A. D., 1996, Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, 81, 660–679.

[33]. Ones, D. S., Viswesvaran, C., & Schmidt, F. L., 1993, Comprehensive meta-analysis of integrity test validities. *Journal of Applied Psychology*, 78, 679–703.

[34]. Raven, J., Raven, J. C., & Court, J. C., 1998, Manual for the Raven's Progressive Matrices and Vocabulary Scales. *Oxford: Oxford Psychologists Press.*

[35]. Roberts, B. W., Chernyshenko, O. S., Stark, S. E., & Goldberg, L. R., 2005, The structure of conscientiousness: An empirical investigation. *Personnel Psychology*, 58, 103–139.

[36]. Schmidt, F. L., & Hunter, J. E., 1998, The validity and utility of selection methods in personnel psychology. *Psychological Bulletin*, 124, 262–274.

[37]. Schmit, M. J., Kihm, J. A., & Robie, C., 2000, Development of a global measure of personality. *Personnel Psychology*, 53, 153–193.

[38]. Scholz, G., & Schuler, H., 1993, Das nomologische Netzwerk des Assessment Centers: Eine Meta analyse. Zeitschrift für Arbeits- und Organisations psychologie, 37, 73–85.

[39]. Viswesvaran, C., & Ones, D. S., 2000, Measurement error in Big Five personality assessment. *Educational and Psychological Measurement*, 60, 224–235.

[40]. Watson, G. B., & Glaser, E. M., 1980, Watson–Glaser Critical Thinking Appraisal Manual. *San Antonio,*

TX: Psychological Corporation.
[41]. Wesman, A. G., 1965, Wesman Personnel Classification Test Manual. *New York: Psychological Corporation.*