Human Performance

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/hhup20

The Expert Assessor Perspective on Assessment Center Taxonomies

Duncan J. R. Jackson, George Michaelides, Chris Dewberry & Wei-Ning Yang

To cite this article: Duncan J. R. Jackson, George Michaelides, Chris Dewberry & Wei-Ning Yang (25 Nov 2024): The Expert Assessor Perspective on Assessment Center Taxonomies, Human Performance, DOI: 10.1080/08959285.2024.2428190

To link to this article: https://doi.org/10.1080/08959285.2024.2428190

HUMAN PERFORMANCE
https://doi.org/10.1080/08959285.2024.2428190

Routledge
Taylor & Francis Group

The Expert Assessor Perspective on Assessment Center Taxonomies

Duncan J. R. Jackson[a], George Michaelides[b], Chris Dewberry[c], and Wei-Ning Yang[a]

[a]King's College London; [b]University of East Anglia; [c]Independent Scholar

ABSTRACT

Research on assessment centers (ACs) has advanced the development of taxonomies for the evaluation of dimension and exercise performance. However, largely missing from current AC taxonomies is the perspective of subject matter expert assessors. Assessor perspectives could contribute toward improving construct differentiation and the development of a theoretical understanding of AC dimensions and exercises. In this study, 197 internationally based assessors participated in a series of multidimensional scaling (MDS) tasks involving stimuli from extant dimension and exercise taxonomies. For our dimension taxonomy, results suggested distinctions between (a) task orientation, (b) interpersonal relations, (c) activity, (d) organizing and planning, and (e) tolerance for stress/uncertainty. For our exercise taxonomy, results suggested that exercises are distinguished by varying levels of (a) media richness and (b) interpersonal interaction. We investigated assessor perceptions of illustrative dimension-exercise combinations and found, in contrast to findings for operational ACs, that assessors expected to rate same dimensions across different exercises. Our results suggest updates to and unique perspectives on taxonomies for dimensions and exercises. We discuss novel theoretical and practical insights that contribute to knowledge relevant to AC research and practice. We furthermore offer applied methodological contributions based on our unique application of MDS in the context of assessor perceptions.

Assessment centers (ACs) are a multifaceted evaluation approach used to guide selection and employee development decisions (Sackett et al., 2017; Speer, Christiansen, Goffin, et al., 2014). ACs often involve the evaluation of individuals by trained assessors on work-related dimensions (e.g., communication skills, teamwork) across two or more work simulation exercises (e.g., group discussions, role plays, Kleinmann & Ingold, 2019). Despite the widespread use of ACs for high-stakes decision-making, their measurement properties have long been debated (Lievens & Christiansen, 2012). Although they represent the most common scoring basis for ACs, dimensions tend to explain only small portions of variance in AC ratings (Lievens, 2009). Debates relevant to this finding have led to alternative scoring approaches for ACs involving exercise rather than dimension scores (Lance, 2012).

The identification of measurement problems in ACs, particularly around whether ACs measure dimension or exercise scores (e.g., Lievens, 2009; Sackett & Dreher, 1982), has motivated two strands of research: one into dimensional taxonomies (Arthur et al., 2003; Meriac et al., 2014) and the other into exercise taxonomies (Hoffman et al., 2015). Relative to other areas of scrutiny (e.g., personality and cognitive ability, Barrick et al., 2001; Neisser et al., 1996), a priori studies on construct development in ACs remain scarce. Yet, the limited body of

CONTACT Duncan J. R. Jackson duncan.jackson@kcl.ac.uk King's Business School, King's College London, Bush House, 30 Aldwych, London WC2B 4BG, England

Supplemental data for this article can be accessed online at https://doi.org/10.1080/08959285.2024.2428190

literature on this topic has provided encouraging evidence in support of AC taxonomies. In general terms, research on taxonomies is of substantial value to industrial-organizational (I-O) psychology, as evidenced by its prevalent role in the literature on performance evaluation (Borman & Brush, 1993), feedback management behavior (Moss et al., 2003), and adaptive performance (Pulakos et al., 2000).

While the few studies that exist on AC taxonomies offer promise, their development primarily involved a process that was (a) part of a tangential step toward some other primary aim (usually conducting a meta-analysis) rather than a primary research focus and, likely as a consequence, (b) were based on perceptions from small numbers of subject matter experts (SMEs, Arthur et al., 2003, $N = 9$; Hoffman et al., 2015, minimum $N = 2$; Meriac et al., 2014, $N = 3$), who were primarily either doctoral students or academic researchers. We acknowledge that doctoral students and academic researchers may hold experience as AC assessors. However, previous studies lack information relating to SME background experience and included small SME samples in their taxonomy development process.

In the present study, we contribute a novel approach to the development of AC taxonomies by focusing on SME conceptual differentiations between different dimensions, exercises, and their combination. We approach this contribution from the unique perspective of a relatively large group of SMEs who hold considerable experience in assessing AC participants. In ACs, the SME perspective is typically provided by trained and experienced assessors who hold operational knowledge about how dimensions and exercises are applied in ACs (Dewberry & Jackson, 2016). This perspective could assist in the development and refinement of taxonomies that inform on the latent structure of dimensions and exercises. This idea moreover addresses calls in the literature for the development of a theoretical understanding about the variables intended for evaluation in ACs (Anderson et al., 2006; Kolk et al., 2004; Merkulova et al., 2016).

We present three main contributions to the AC literature in the present study. First, in the frame of the only study we could identify where the primary focus is on taxonomy development, we contribute the perspective of a relatively large number of SME assessors to AC taxonomies. Second, we provide a unique viewpoint on SME perceptions about the latent structure of dimensions, exercises, and dimension-exercise combinations with a view to developing theory on AC measurement. Third, we provide a methodological contribution in our application of multidimensional scaling (MDS, see Hair et al., 1998; Meyers et al., 2013) as an approach toward capturing SME judgments in the development of taxonomies in I-O psychology. To our knowledge, MDS has not yet been applied to SME judgments in the context of ACs.

Development of a dimension taxonomy

Dimensions typically form the main scoring basis for ACs, and a sizable literature has emerged on research relating to their structure (Borman, 2012). Some of the earliest developmental work on AC dimension taxonomies can be traced to an exploratory factor analysis (EFA) of averaged dimension composites from 133 Office of Strategic Services (OSS) assessees (Handler, 2001; The OSS Assessment Staff, 1948). Five decades after this study, Arthur et al. (2003) pioneered the development of a nonmilitary taxonomy for AC dimensions.

As a starting point for their taxonomy, Arthur et al. (2003) referred to the 33 "Common Managerial Dimensions" listed in Thornton and Byham (1982), pp. 138–140). They sought out representative studies from the research literature and identified 168 different dimension titles from a range of AC studies. With the aim of data reduction, the researchers engaged in a systematic coding exercise involving a team of six I-O psychology graduate students together with two Arthur et al. authors and one subject matter expert (see p. 132). The seven dimensions identified from this process included: consideration/awareness of others, communication, drive, influencing others, organizing and planning, problem solving, and tolerance for stress/uncertainty. While Arthur et al. cautioned theirs was

a "first attempt" (p. 148) to develop a taxonomy for dimensions, it has nonetheless been widely adopted in the AC literature (e.g., Bowler & Woehr, 2006; Dilchert & Ones, 2009; Meriac et al., 2008).

Refinement of the Arthur et al. taxonomy and its theoretical underpinnings was called for by Meriac et al. (2014). To provide a theoretical grounding, Meriac et al. reviewed up-to-date background literature relevant to dimensions. In turn, they applied confirmatory factor analysis (CFA) to inform on the fit of ratings from existing ACs to theoretical propositions identified in their literature review. To develop their taxonomy, the authors utilized a coding process involving two authors of their article with the assistance of a doctoral student in I-O psychology. Meriac, et al. reviewed several theoretical perspectives to inform on the results of their summary dimension taxonomy. They settled on a 3-factor solution, including: administrative skills, relational skills, and drive (based on, for example, Borman & Motowidlo, 1997; Thornton & Byham, 1982).

Development of an AC exercise taxonomy

In response to criticisms related to the common observation that only small portions of variance in operational AC ratings are associated with dimensions (e.g., Lievens & Christiansen, 2012), alternative, exercise-based scoring approaches have been suggested for ACs. Dating back over three decades (Goodge, 1988; Lowry, 1997), exercises or "tasks" have been proposed and applied in practice as a scoring foundation for ACs (Jackson et al., 2011; Lance, 2012; Thoresen & Thoresen, 2012). Given the number of exercises that could potentially be used in an AC, the application of exercise scores raises questions about whether a summary taxonomy could be developed for exercise type.

This proposition was addressed by Hoffman et al. (2015), who, via a method similar to that applied in Arthur et al. (2003), coded exercises from studies of ACs into a summary list. As a basis for their coding approach, Hoffman et al. referred to a taxonomy of five common exercises summarized in Thornton (1992). The original Thornton exercise taxonomy was itself based on findings from a practice survey (Gaugler et al., 1990), and included nine exercise types (discussed below). Variations on the Thornton taxonomy have been applied elsewhere in AC research (e.g., Bowler & Woehr, 2006; Lievens et al., 2006). In the present study, we primarily focus on the most recent, five-exercise version provided in Hoffman et al. (2015), which includes in-baskets, leaderless group discussions (LGDs), role plays, case analyses, and oral presentations.

Hoffman et al. (2015) included 69 published and unpublished studies of AC exercises for coding into their five-exercise taxonomy. In their classification process, two or more coders were involved: at least two of whom were graduate students for any given coding task (C. Lance, personal communication, September 8, 2021). With coder agreement at 86%, the Hoffman et al. (2015) study provided encouraging evidence for the Thornton-based exercise taxonomy. Nonetheless, it could be that the exercise categories from the Thornton taxonomy form meaningfully related subclusters. Specifically, and according to a currently unidentified basis, some exercises might be perceived by SMEs as conceptually similar, whilst others might be perceived as distinct.

Construct differentiation

The extant dimension and exercise taxonomies described above have been widely adopted in the discipline (e.g., Bowler & Woehr, 2006; Dilchert & Ones, 2009; Lievens et al., 2006). Nonetheless, considerations relevant to AC taxonomies remain unresolved, including those relating to construct differentiation and the theoretical basis for dimension and exercise subclusters.

One of the problems routinely identified in ACs is that dimensions tend to be poorly differentiated from one another. A clear representation of this effect can be found in Bowler and Woehr (2006), where the authors reported an average meta-analytic correlation between dimensions of .79. This finding is possibly suggestive of an undifferentiated general performance factor that manifests across exercises. General factors have been identified as a key component of AC measurement structure in previous literature (Jackson et al., 2016; Lance et al., 2007; Putka & Hoffman, 2013). In their highly

controlled study, Putka and Hoffman estimated the proportion of dimension variance in AC ratings at around 3%, whereas general performance was estimated at around 46%. Similarly, Jackson et al. estimated dimension variance at around 1% and general performance at around 54%.

Previous research has suggested approaches toward improving conceptual differentiations between constructs evaluated in ACs. In Heimann et al. (2022), traditional AC dimensions were replaced with Big Five Model (BFM) traits. Personality traits have been the subject of extensive background research and refinement to an extent exceeding that for AC dimensions (e.g., Barrick & Mount, 2005). Across two samples, Heimann et al. found that BFM trait effects explained 25% and 33%, respectively, of the variance in AC ratings. The average factor correlation[1] estimated between different BFM traits for Sample 1 = .34 (*SD* = .19) and for Sample 2 = .56 (*SD* = .25). These results suggest an improvement vis-à-vis construct differentiation when compared to the .79 average factor correlation meta-analytic estimate from Bowler and Woehr (2006). A possible reason for this improvement might be that it is easier to differentiate between BFM traits than dimensions.

However, performance dimensions, rather than personality traits, are usually applied in operational ACs (e.g., Eurich et al., 2009; Krause et al., 2011). The continuation of this practice appears likely, given the perceived job relevance of dimensions, their potential applicability to behavior in job simulations, and their links to competency models (Arthur, 2012; Meriac et al., 2014; Schippmann et al., 2000). Further research into dimensions thus appears justified, given their possible continued use, the low dimension differentiation observed in studies of AC ratings, and the possibility that greater conceptual differentiation between constructs leads to improved construct measurement.

The same logic described above for dimensions applies to the scoring of exercise constructs (see Lievens, 2008) and thus to exercise taxonomies. Speer, Christiansen, Goffin, et al. (2014) found improvements in criterion-related validity for ACs where exercises were differentiated from one another. This finding is intuitive, because well differentiated predictors are more likely to explain unique portions of variance in an outcome (e.g., Spicer, 2005). The findings of Speer, et al. nonetheless raise questions about the basis for deciding how and why AC exercises are differentiated.

The exercise taxonomy based on Thornton (1992) that features in multiple AC publications (e.g., Hoffman et al., 2015; Lievens et al., 2006) covers a substantial range of different exercises. However, to our knowledge, little is known about whether or on what basis any commonalities or distinctions exist between subsets of exercises in this taxonomy. Identification of such patterns could help to explain the theoretical basis for AC exercises and could form criteria on which exercises may be included as distinct predictors in ACs.

Theoretical foundations

On theoretical foundations, Arthur et al. (2003) did not provide theoretical links for their dimension taxonomy, possibly because their chief aim was to reduce data for the practical purpose of conducting a meta-analysis.[2] Meriac et al. directly addressed this issue and suggested theoretical bases for their dimension taxonomy covering administrative skills, relational skills, and drive. Administrative skills were conceptually likened to general mental ability (GMA) and relational skills and drive were likened to the traits "getting along" and "getting ahead," respectively (see R. Hogan & Kaiser, 2005). Getting along and getting ahead are analogous to the consideration and initiating structure constructs from the Ohio State Leadership Studies (Stogdill, 1950). Where consideration is concerned with interpersonal relations, initiating structure is concerned with task orientation (Judge et al., 2004).[3] Despite developments arising from these findings, dimension measurement has continued to present a challenge for

[1]Average factor correlations were estimated for the present study based on online supplementary material linked in the Heimann et al. (2022) article.

[2]We note this as an observation rather than a criticism. Even popular models of personality have roots that are atheoretical (e.g., Laher, 2013).

[3]The conclusions in Meriac et al. (2014) were similar to those in the Thornton and Byham (1982) review of factor analytic evidence for dimensions (i.e., the summary dimensions in Thornton and Byham were: administrative skills, interpersonal skills, and activity).

ACs (Jackson et al., 2022). This was even the case in the Putka and Hoffman (2013) AC where dimensions were evaluated that were similar to those identified in the Meriac et al. (2014) taxonomy (e.g., knowledge of administrative procedures; relating to others).

While at least some background theoretical development is available for dimensions, far less is available for exercises. Our review of the limited body of literature on the basis for exercises suggests that distinctions between exercises in the Thornton (1992) taxonomy could possibly concern (a) behavioral output and (b) social interaction. Depending on their content and design, different exercises require distinct types of behavioral output (Speer, Christiansen, Melchers, et al., 2014). To illustrate, LGDs or presentations often require verbal responses. In contrast, in-baskets often require written responses. A similar distinction between verbal and written output is discussed in the literature on consumer behavior. Maity et al. (2018) refer to *media richness*, where sources of information can be defined as rich (including visual and auditory components) versus lean (including mainly text). This distinction could be relevant to the classification of AC exercises.

On distinctions between exercises involving levels of social interaction, Dierdorff et al. (2012) suggest that features of the social context (e.g., the degree to which interdependency affects role performance) determine the importance of social interactions and their maintenance at work. Hoffman et al. (2015) found evidence supporting this idea in their research on exercise characteristics. Similarly, both Lance (2008) and Jackson (2012) discuss the potential for AC exercises to be conceptualized as simulations of role behavior. In this view, an AC exercise is considered as a measure of a job-relevant role that is defined by specific types of interpersonal interaction. For example, a discussion exercise designed to evaluate a supportive managerial role might require a high level of interpersonal interaction. Knowledge about how exercise subsets are distinguished from one another could inform on how this role-based perspective relates to levels of social interaction.

An expert assessor perspective

Coders involved in the development of AC taxonomies to date were, as discussed above, primarily graduate students or academic researchers with some minority input from SMEs (see Arthur et al., 2003; Hoffman et al., 2015; Meriac et al., 2014). However, it is trained and experienced assessors who observe responses, provide an evaluation of those responses, and classify them into summary frameworks (Lievens, 2001). Assessors are often rigorously trained on rating and scoring procedures using standard-setting approaches (e.g., frame-of-reference training, Roch & O'Sullivan, 2003).

Once they have gained a sufficient degree of training and experience, assessors may be regarded as AC SMEs akin to those in related fields (e.g., Tross & Maurer, 2000). Because of their experiential knowledge (e.g., Schmidt & Hunter, 1993), it is likely that assessors hold expertise not only about AC design features, but also about how those features interact and operate. Assessor SMEs thus appear well positioned to inform on the conceptual refinement of AC taxonomies for dimensions and exercises.

Summary and research questions

Despite a lengthy history underpinning AC research and practice, it is only in the last two decades that taxonomies have been developed for AC dimensions and exercises (Arthur et al., 2003; Hoffman et al., 2015; Meriac et al., 2014). It is suggested in the current literature that further refinement of dimension taxonomies is justified because (a) when rated, dimensions are often found to be poorly differentiated (Bowler & Woehr, 2006; Jackson et al., 2016; Putka & Hoffman, 2013) and (b) enhanced conceptual distinctions between constructs possibly leads to improved construct differentiation (Heimann et al., 2022). Regarding exercises, almost no literature exists on the development of an exercise taxonomy beyond early work by Thornton (1992) and Gaugler et al. (1990). Thus, further development of exercise taxonomies is required to contribute to an understanding about the basis for exercises and exercise-based scoring in ACs.

Further development of taxonomies for ACs could lead to an improved theoretical understanding about the basis for dimensions and exercises and could be applied in future research endeavors (e.g., in meta-analyses). Enhanced taxonomies could facilitate the practical development of ACs, where a key aim is to generate meaningfully differentiated scores. Well-differentiated scores could be used to enhance criterion-related validity in employee selection or to provide meaningful feedback in employee development. Trained and experienced assessors, although currently underrepresented in the development of AC taxonomies, are well placed to inform on AC development, given their relevant background expertise. We address our general aim of developing AC taxonomies from the perspective of the MDS analytical framework (we provide detail about MDS below). This presents a methodological contribution because the unique perspective offered by MDS has not, to our knowledge, been previously applied in the AC context. The research gaps and aims we describe above lead to our first two Research Questions (RQs), as follows:

RQ1: How do perceived distances between different AC dimensions indicate an expert cognitive representation of a dimension taxonomy?

RQ2: How do perceived distances between different AC exercises indicate an expert cognitive representation of an exercise taxonomy?

In addition to considerations relating to the development of dimension and exercise taxonomies, the AC measurement design raises questions about how taxonomies are applied. Dimensions constitute the most common scoring approach in ACs (Krause & Thornton, 2009; Krause et al., 2011). However, assessors typically score dimensions across exercises and thus focus on dimension-exercise combinations (Thornton & Byham, 1982). To foster an understanding about the application of AC taxonomies, we investigate patterns relating to dimension-exercise combinations. We note this aim does not involve comprehensively accounting for perceptions relating to the combination of all possible dimensions with all possible exercises. Such an undertaking would likely raise practical restrictions. Rather, our aim here is to facilitate an understanding of assessor perceptions relating to an illustrative and relevant subset of AC dimensions and exercises. This leads to our third RQ as follows:

RQ3: How do perceived distances between illustrative AC dimension-exercise combinations indicate an expert cognitive representation of how dimensions are applied in exercises?

Method

Transparency and openness

Below, we describe our sampling plan, relevant sample information, and the measures used in our study. We adhered to APA journal article reporting standards for quantitative research (see Appelbaum et al., 2018). All data, analysis code, output, and supplementary materials are available at https://osf.io/gvqux/?view_only=d278110d541a42debd710409100f11ae. All analyses were conducted using IBM SPSS Statistics Version 29 (IBM Corp, 2023). The present study design and its analysis were not preregistered.

Participants

Participants included 197 AC assessors (118 women, 78 men, and 1 other), with a mean age of 32.29 ($SD = 12.18$) years. Participation in this study was voluntary and informed consent was obtained. Regarding sampling and recruitment, potential participants were sent a link to an online survey via various online forums relevant to ACs, human resources, and I-O psychology primarily

via LinkedIn Groups. Special interest groups were also contacted, and links were sent out via member networks or were posted on their LinkedIn pages (e.g., the United Kingdom Assessment Centre Group, Gateway I-O Psychologists Group USA, Chartered Institute of Personnel and Development). As detailed in Table 1, most participants were White (85%) with just over half originating from the UK (54%) with several smaller percentages (<11% each) representing multiple international locations.

Most participants (66%) reported holding an academic qualification in I-O psychology and around 69% reported holding a postgraduate qualification. All participants were trained and experienced in the application of ACs. Participants reported assessing a median of 30 ACs (interquartile range, IQR = 192 ACs). Participants reported receiving a median of 7 days' AC training (IQR = 9 days). Typical training content and coverage for the assessors is presented in Table 1.

Table 1. Assessor demographic characteristics and background.

Variable	%
Gender	
Men	39.59
Women	59.89
Other	.51
Racial background	
Black	4.60
White	84.77
Asian	7.61
Other	3.00
Region	
United Kingdom	53.77
Greater Europe	10.55
North America	9.05
Australasia	10.05
Asia	6.53
Other	10.05
Education	
Completed high or secondary school	7.50
Bachelor's degree	11.00
Postgraduate diploma	8.50
Master's degree	56.00
PhD	13.00
Other	4.00
Multiple response: Qualifications[a]	
Professional HRM-related qualification (e.g., CIPD, SHRM)	28.80
Academic HRM-related qualification (e.g., HRM, HRD)	19.19
Academic I-O qualification	65.70
Professional qualification unrelated to HRM or I-O	14.14
Other qualification unrelated to HRM or I-O	15.15
No qualifications, experience with ACs only	6.60
Other	5.60
Multiple response: Training coverage[a]	
Common rater errors (e.g., like-me bias, halo effects)	88.42
Frame-of-reference training	52.63
Observing, rating, and recording behavior	95.26
Exercise content	87.89
Rating performance on exercises	87.89
Rating dimensions within exercises	87.37
Rating dimensions across exercises	70.00
Rating mock assessees	67.37
Other	11.58

Total $N = 197$ (118 women, 78 men, 1 other). [a] Multiple response items shown as percent of cases (i.e., percent of people responding yes to each item). HRM = human resource management, CIPD = Chartered Institute for Personnel and Development, SHRM = Society for HRM, HRD = human resource development, I-O = industrial and organizational psychology or occupational psychology, AC = assessment center.

MDS tasks

The bulk of our survey consisted of a series of MDS tasks (see Meyers et al., 2013; Nunnally & Bernstein, 1994). MDS presents pairs of stimuli to participants, who are required to indicate the degree to which they perceive similarities between each stimulus pair. This process is repeated until every possible pairing combination is exhausted for the set of stimuli under scrutiny. Similarity ratings are typically reverse-scored and then averaged in MDS, resulting in a dissimilarity matrix that summarizes the degree of separation between each stimulus pairing in Cartesian space (Hair et al., 1998). Higher values in a dissimilarity matrix thus indicate greater perceived separation between stimuli, whereas lower values indicate closer perceived proximity between stimuli. Dissimilarity matrices are used as data input for an MDS analysis. The output from MDS is a visual representation of the relationships between each stimulus. Variables that cluster together more closely in this representation are perceived as related. Variables that are relatively distant from one another are perceived as less related (Young & Harris, 2012).

Hair et al. (1998) contrast two common outcomes in MDS, including the identification of (a) previously unrecognized continua, versus (b) "comparative evaluations" between stimuli (p. 527). Whether it is possible to draw conclusions about a perceived continuum (i.e., x and y are separated, but belong to the same continuum) or a comparative evaluation (i.e., a simple separation between constructs, where x and y are conceptually different from one another) from an MDS analysis depends on the study design, the variables under scrutiny, and the study outcomes. Continua are identified where two stimuli are differentiated such that they refer to two extremes that belong to the same element[4] (to illustrate, pole 1 = light green, pole 2 = dark green, common element = the color green). Comparative evaluations are identified where two stimuli are differentiated because they are perceived as conceptually different, but not because they exist as poles on a continuum (e.g., pole 1 = color, pole 2 = taste, which identifies two separate constructs that are not assumed to exist on a common element).

Regarding the analyses that follow, no dimension taxonomy in previous literature identifies a perceived continuum linking different dimensions. To our knowledge, in all cases, the aim in previous research has been to identify dimensions that are conceptually distinct (e.g., Arthur et al., 2003). Our expectations for the outcomes of our dimension-related MDS tasks align with those in previous research around identifying distinct dimensions. In MDS terms, we anticipate identifying evidence for a comparative evaluation of dimensions.

In contrast, and due to a lack of current research, our expectations regarding exercises are more flexible. In the only known study on AC exercise taxonomies (Hoffman et al., 2015), coders arranged exercises from AC research into categories, which implies a comparative evaluation. However, it is possible that previously unrecognized continua may be relevant to AC exercises (e.g., degree of interpersonal interaction, ranging from none in a written exercise to a high degree in a group exercise). Given the previous findings of Hoffman et al. suggesting comparative evaluations and the contrasting possibility of continua, we leave open the possibility of either outcome in the present study.

Dimension and exercise stimuli

The list of dimensions included as stimuli in our MDS task (see RQ1 and Table 2) was based on the seven summary dimensions developed by Arthur et al. (2003). We retained the dimension tolerance for stress/uncertainty in our list given its inclusion in the original Thornton and Byham (1982) list and its potential relevance in contemporary organizational contexts (e.g., in studies of resilience, see Blustein et al., 2022; Rolin et al., 2022). The exercises included as stimuli in our MDS task included the list of five summary exercises (see RQ2 and Table 3) from Hoffman et al. (2015). Definitions for

[4]We opted for the term *element* rather than the traditional MDS term *dimension* or *dimensional distinction* to avoid confusion with AC dimensions.

Table 2. Dimension titles and definitions included as initial multidimensional scaling stimuli.

Dimension title	Definition
Communication	The extent to which an individual conveys oral and written information and responds to questions and challenges
Consideration and awareness of others	The extent to which an individual's actions reflect a consideration for the feelings and needs of others as well as an awareness of the impact and implications of decisions relevant to other components both inside and outside the organization.
Organizing and planning	The extent to which an individual systematically arranges his or her own work and resources as well as that of others for efficient task accomplishment and anticipates and prepares for the future.
Problem solving	The extent to which an individual gathers information; understands relevant technical and professional information; effectively analyzes data and information; generates viable options, ideas, and solutions; selects supportable courses of action for problems and situations; uses available resources in new ways; and generates and recognizes imaginative solutions.
Influencing others	The extent to which an individual persuades others to do something or adopt a point of view in order to produce desired results and takes action in which the dominant influence is the individual's own convictions rather than the influence of others' opinions.
Drive	The extent to which an individual generates and maintains a high activity level, sets high performance standards and persists in their achievement, and expresses the desire to advance to higher job levels.
Tolerance for stress/ uncertainty	The extent to which an individual maintains effectiveness in diverse situations under varying degrees of pressure, opposition, and disappointment.

Definitions in this table were adapted from Thornton and Byham (1982), Meriac et al. (2014), and Arthur et al. (2003).

Table 3. Exercise titles and definitions included as initial multidimensional scaling stimuli.

Exercise title	Definition
In-baskets	An individual activity that simulates administrative work including a simulated set of memos, messages, e-mails, letters, and reports, such as those that might accumulate in a manager's "in-basket," as well as other reference material (e.g., organizational charts, personal calendars). The materials are usually interrelated and vary with respect to complexity and urgency. Participants are typically asked to play the role of a person new to the job, working alone with the goal of trying to clear the in-basket.
Leaderless group discussions	A group activity typically requiring four to eight participants to solve a problem within a time limit. This might involve a written solution to be agreed by all members of the group. Specific roles may be assigned to the various group members. However, no one is assigned the role of leader or chair. Rather, leadership behaviors are allowed to emerge during the discussion.
Role plays	A simulation in which a participant responds to a scripted scenario presented by an actor (e.g., a dissatisfied customer or an employee with a grievance).
Case analyses	An activity where multiple participants are provided with a description of a complex problem or set of problems, often relating to a management scenario. The group is required to discuss and potentially agree on solutions to the problem(s) described.
Oral presentations	An activity where information is reported orally, often by an individual, to an audience. The presentation might involve reference to presentation slides or other stimuli to facilitate understanding.

Definitions in this table were adapted from Thornton (1992) and Hoffman et al. (2015).

our exercise list were further informed by other sources in the extant literature (Cook, 2016; Gatewood et al., 2016; Greenhaus & Callanan, 2006; Spychalski et al., 1997; Thornton & Byham).

Most of the exercise definitions we reviewed were largely uncomplicated, with the possible exception, at least in part, of the case analysis. For example, in the description provided by Thornton and Byham (1982), the stimulus in the case analysis involves an examination of a written description of a management problem. However, the mode of presentation following this examination is described as either written, verbal in groups of assessees, or verbal between individual assessees and assessors. To maintain simplicity and to better reflect a range of different exercise formats, we focused on the written description format for the case analysis coupled with a group-based mode of presentation.

To develop a shared schema for classification related to the MDS tasks in our study, participants were provided with definitions of both dimensions and exercises (see Tables 2 and 3) and were required to peruse these definitions prior to completing their MDS tasks.

Same dimension, different exercise expectations

We requested that participants provide their perspective on ratings of same dimensions across different exercises to inform on assessor perceptions of dimension-exercise combinations (see RQ3). The basis of this aim was to address expectations when assessors are presented with an illustrative and relevant set of dimensions intended for evaluation across different exercises. Following this idea, we retained three dimensions that were more likely to be conceptually distinct according to definitions provided in Arthur et al. (2003) and three exercises that took different formats according to descriptions provided in Thornton (1992). The three dimensions we retained were communication, problem solving, and drive. The three exercises we retained were the in-basket, role play, and oral presentation.

Analytic approach

For all RQs, we applied classic metric MDS (see Knezek et al., 2023; Tziner et al., 2020; Vera & Macias, 2021). Initially, participants were presented with all possible pairs of the dimensions, exercises, and dimension-exercise combinations described above. Participants were required to rate the degree of similarity between variables in each pairing on a scale from 1 (extremely dissimilar) to 5 (extremely similar). We reverse-scored these similarity ratings to produce average dissimilarity matrices that would serve as data input (see Meyers et al., 2013; Nunnally & Bernstein, 1994). The MDS procedure in IBM SPSS applies the alternating least squares scaling algorithm (or ALSCAL) originally developed by Takane et al. (1977). ALSCAL adopts a least squares approach to estimating coordinates for each stimulus and represents these coordinates in geometric space (Hair et al., 1998; Meyers et al., 2013; Vera & Macias, 2021). The distances between any coordinates are thus intended to represent the original averaged dissimilarity judgments of respondents.

We employed commonly applied MDS statistical indicators to provide guidance on the appropriate number of elements likely to be present in each dissimilarity matrix (e.g., Young & Harris, 2012). These indicators included Young's S-stress index, Kruskal's stress index, and *R* square estimates. Meyers et al. (2013) provide guidance for the interpretation of these indicators, where stress values $< .05$ = excellent, $< .10$ = good, $< .20$ = fair, $> .20$ = poor fit and for *R* square where values $> .60$ are considered acceptable. Upon arriving at an appropriate number of elements for each matrix, we visually inspected MDS plots to facilitate our interpretation process. Analyses containing ≥ 3 elements are presented in the IBM SPSS Statistics software as output that can be rotated in 3D space. Where relevant, we present multiple angles from this output to aid interpretation for the reader.

Results

The output from an MDS analysis involves a visual representation of the degree of distance between coordinates presented in multidimensional space (Hair et al., 1998). In some cases, the interpretation of this information necessitates the physical rotation of output so that elements can be compared with one another. Where relevant (i.e., where a 3-D representation is required), we have attempted to facilitate this process by presenting our MDS output from different angles. We recognize that MDS is not commonly applied in I-O psychology, and thus we provide guidance on the interpretation of our results which can be accessed via our supplemental material. Notably, this supplemental material includes (a) different angles of output with overlayed ellipses representing patterns of interest and (b) original IBM SPSS output that can, where relevant and if the reader has access to the IBM SPSS program, be rotated in 3-D.

Table 4. Dissimilarity matrix for assessment center dimensions.

	Assessment Center Dimension *M*s	1	2	3	4	5	6
1	Communication						
2	Consideration	2.52					
3	Organization and planning	3.73	3.48				
4	Problem solving	3.70	3.46	2.29			
5	Influencing others	1.92	2.01	3.66	3.25		
6	Drive	4.05	3.86	3.27	3.41	3.39	
7	Tolerance for stress/uncertainty	3.89	3.51	3.32	3.22	3.59	3.18
	Assessment Center Dimension *SD*s						
1	Communication						
2	Consideration	1.15					
3	Organization and planning	1.24	1.30				
4	Problem solving	1.27	1.30	1.09			
5	Influencing others	.98	.91	1.25	1.26		
6	Drive	1.06	1.18	1.25	1.30	1.30	
7	Tolerance for stress/uncertainty	1.14	1.28	1.21	1.31	1.21	1.25

Average mean (*M*) distances are presented in the upper matrix and matching standard deviations (*SD*s) are provided in the lower matrix. The median *SD* = 1.25. Scale points ranged from 1 (extremely similar) to 5 (extremely dissimilar).

MDS structure: dimensions, exercises, and combinations

Tables 4 through 6 show dissimilarity matrices for analyses relating to dimensions, exercises, and dimension-exercise combinations, respectively (including both *M* and corresponding *SD* values).[5] As we reverse-scored SME responses, values in Tables 4 through 6 are interpreted on a scale ranging from 1 (extremely similar) to 5 (extremely dissimilar). Table 4 suggests variation on average relating to the degree that dimensions were perceived as similar or dissimilar. Influencing others and communication were rated as relatively similar on average ($M = 1.92$). Whereas drive and communication were rated as relatively dissimilar ($M = 4.05$). Table 5 also suggests variability in the perceived similarity between exercises. LGDs were perceived as generally different from in-baskets ($M = 4.38$), whereas case analyses and in-baskets were perceived as relatively similar ($M = 2.47$). Table 6 moreover suggests variability regarding dimension-exercise combinations. Nonetheless, raw dissimilarity matrices are subject to statistical conflation. Their main purpose is generally as data input into a controlled MDS analysis.

Table 5. Dissimilarity matrix for assessment center exercises.

	Assessment Center Exercises *M*s	1	2	3	4
1	In-baskets				
2	Leaderless group discussions	4.38			
3	Role plays	3.96	2.93		
4	Case analyses	2.47	3.42	3.72	
5	Oral presentations	3.92	3.15	3.04	3.43
	Assessment Center Exercises *SD*s				
1	In-baskets				
2	Leaderless group discussions	.92			
3	Role plays	1.13	1.26		
4	Case analyses	1.14	1.24	1.20	
5	Oral presentations	1.18	1.20	1.19	1.16

Average mean (*M*) distances are presented in the upper matrix and matching standard deviations (*SD*s) are provided in the lower matrix. The median *SD* = 1.19. Scale points ranged from 1 (extremely similar) to 5 (extremely dissimilar).

[5]We do not report indicators of interrater agreement because they are of questionable relevance in the context of MDS, which involves the aim of summarizing a distribution of responses. Beyond this aim, there are no expectations around seeking a specific level of agreement among individuals completing MDS tasks.

Table 6. Dissimilarity matrix for assessment center dimension and exercise combinations.

	Dimension/Exercise *M*s	1	2	3	4	5	6	7	8
1	Com-In-basket								
2	Com-Role play	3.57							
3	Com-Oral pres	3.29	2.01						
4	Prob-In-basket	3.21	3.70	3.89					
5	Prob-Role play	3.66	3.05	3.37	2.47				
6	Prob-Oral pres	3.69	3.42	3.08	2.52	2.06			
7	Drive-In-basket	3.62	4.03	3.97	3.39	3.91	3.96		
8	Drive-Role play	3.97	3.27	3.47	3.84	3.32	3.54	2.81	
9	Drive-Oral pres	3.90	3.42	3.19	3.90	3.61	3.35	2.76	2.21
	Dimension/Exercise *SD*s								
1	Com-In-basket								
2	Com-Role play	1.32							
3	Com-Oral pres	1.25	.94						
4	Prob-In-basket	1.36	1.34	1.31					
5	Prob-Role play	1.40	1.36	1.39	1.01				
6	Prob-Oral pres	1.26	1.33	1.34	1.04	0.91			
7	Drive-In-basket	1.30	1.14	1.19	1.23	1.18	1.21		
8	Drive-Role play	1.28	1.32	1.34	1.24	1.31	1.28	1.15	
9	Drive-Oral pres	1.24	1.32	1.32	1.18	1.26	1.24	1.15	.97

Average mean (*M*) distances are presented in the upper matrix and matching standard deviations (*SD*s) are provided in the lower matrix. The median *SD* = 1.26. Scale points ranged from 1 (extremely similar) to 5 (extremely dissimilar). Com = communication, Oral pres = oral presentation, Prob = problem solving.

The MDS analysis commences with a decision about the number of elements each analysis should include. This decision is guided by statistical criteria, as outlined above. The criteria we applied suggested a 3-element solution for AC dimensions (stress = .09, *R* square = .93, for Table 4), a 2-element solution for exercises (stress = .18, *R* square = .78, for Table 5), and a 3-element solution for exercise-AC dimension combinations (stress = .09, *R* square = .94, for Table 6). No other potentially feasible solutions (e.g., a 2-element solution for AC dimensions) for any of our analyses were suggestive of an acceptable fit in our data sets.

Dimension elements

Figure 1 shows the 3-element solution for AC dimensions (see RQ1). The three boxes shown in Figure 1 represent different sides of a cube.[6] To provide guidance for the reader, element 1 is shown in the upper portion of Figure 1 along the x-axis only. When interpreting element 1, any patterns of similarity or separation along the y-axis for element 2 should be ignored. Along the x-axis for element 1 appear drive, tolerance, organizing and planning, and problem solving in one cluster. Separated from this cluster on the x-axis are the dimensions influencing others, consideration and awareness of others, and communication skills, which appear in a separate cluster. Thus, the interpretation of MDS output is based on the idea that assessors perceived dimensions in closer spatial proximity as related and more distant dimensions as separated.

Regarding an interpretation of the dimension solution, element 1 differentiates dimensions concerning task orientation (drive, tolerance, organizing and planning, and problem solving) from those concerning interpersonal relations (influencing others, consideration and awareness of others, and communication skills). Element 2 primarily differentiates organizing and planning and problem solving from drive and tolerance for stress/uncertainty. Element 3 distinguishes tolerance for stress/uncertainty from drive (i.e., a total of five distinct dimensions). None of these perceived differentiations is suggestive of a continuum. Accordingly, and in keeping with previous research, we interpret

[6]A 3-element solution from an MDS analysis does not imply three summary clusters. Rather, it implies that the results of the analysis are necessarily interpreted with reference to a summary object in three spatial dimensions, within which many more distinctions between summary clusters may be relevant.

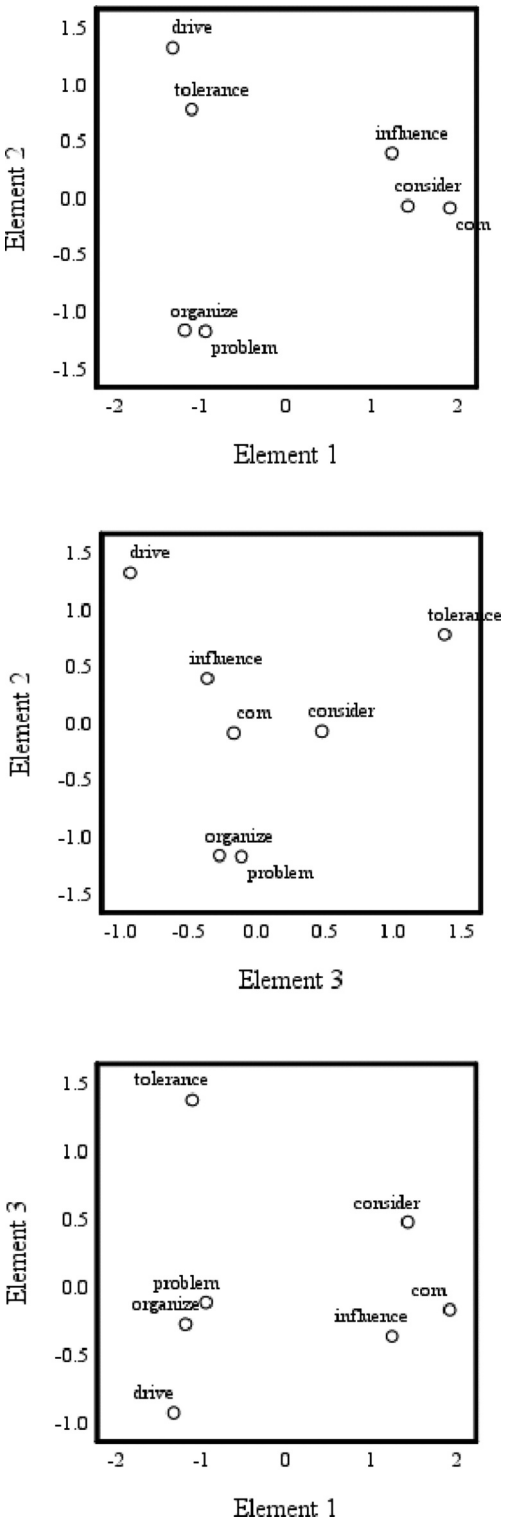

Figure 1. Three-element multidimensional-scaling solution for assessment center dimensions. *Note*. The 3-element solution above is shown from different angles to facilitate interpretation. com = communication skills, organize = organizing and planning, problem = problem solving, consider = consideration and awareness of others, influence = influencing others, tolerance = tolerance for stress/uncertainty. Element 1 differentiates interpersonal from task-oriented dimensions. Element 2 distinguishes dimensions related to preparation from drive. The primary distinction in Element 3 is between tolerance and drive. Given the likelihood of a comparative evaluation in the above, this reflects five distinct dimensions identified in the analysis. Model stress = .09, R^2 = .93.

the MDS elements for dimensions as comparative evaluations (see Hair et al., 1998). Given the pattern of results above in element 1 and the different components of our definition of drive (see Table 2), we suggest that the separation of drive in element 3 possibly indicates activity level. We expand on this idea in our discussion below. Taking all three elements into consideration, the main distinctions suggested in our analysis were between the five dimensions task orientation, interpersonal relations, drive (or activity), organizing and planning, and tolerance for stress/uncertainty.

We note the appearance of the same dimensions in different elements in our MDS solution described above, the cause of which could relate to one of several possibilities (see Meyers et al., 2013 for a discussion on this topic). For example, the role of organizing and planning in element 1 might represent a general capacity to engage with task completion as opposed to maintaining relationships. However, the role of organizing and planning in element 2 may be more specifically about goal-directed resource allocation as distinct from the dimensions drive and tolerance. It is also possible that the reason >1 elements share the same pole in Figure 3 is because of a general lack of conceptual distinctiveness between dimensions.

Exercise elements

As we describe below, and in contrast to the analyses relating to dimensions, our MDS solution for exercises was suggestive of elements that reflected continua. Accordingly, we interpret each element identified in Figure 2 as a perceived continuum in keeping with the suggestions of Hair et al. (1998). To provide evidence in support of this idea, Figure 2 represents the 2-element solution for exercises (see RQ2). Element 1 in Figure 2 differentiated verbal (LGD, role play, and oral presentations) from written data (case analyses and in-baskets), suggestive of a *media richness* element (Maity et al., 2018). Element 2 in Figure 2 differentiated individual (role plays and oral presentations) from group-based interactions (LGDs and case analyses), suggesting an *interpersonal interaction* element (i.e., work roles that differ by degree of interdependency, see Dierdorff et al., 2012). Thus, our analysis for exercises suggests a total of two elements reflective of exercise continua relating to media richness and interpersonal interaction.

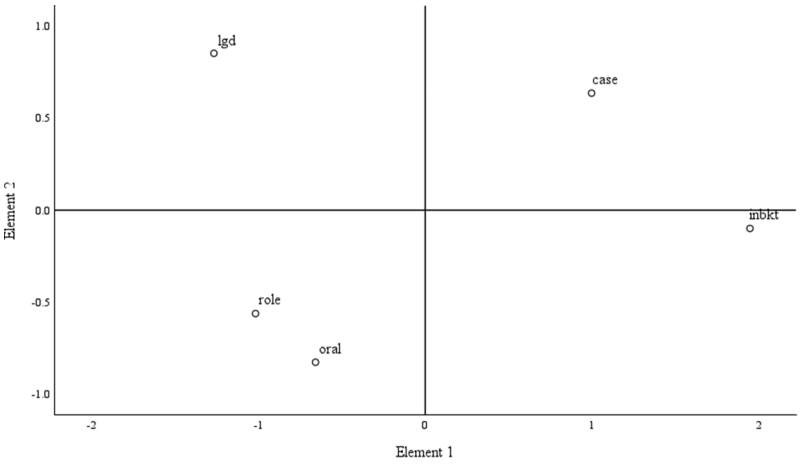

Figure 2. Two-dimensional multidimensional scaling solution for assessment center exercises. *Note*. lgd = leaderless group discussion, case = case analysis, role = role play, inbkt = in-basket, oral = oral presentation. Element 1 primarily differentiates verbal from written sources of information. Element 2 primarily differentiates complexity of social interaction, from group-based interactions to individual interactions. Given the likelihood of a perceived continuum in the above, this reflects two distinct exercise continua identified in the analysis. Model stress = .18, R^2 = .78.

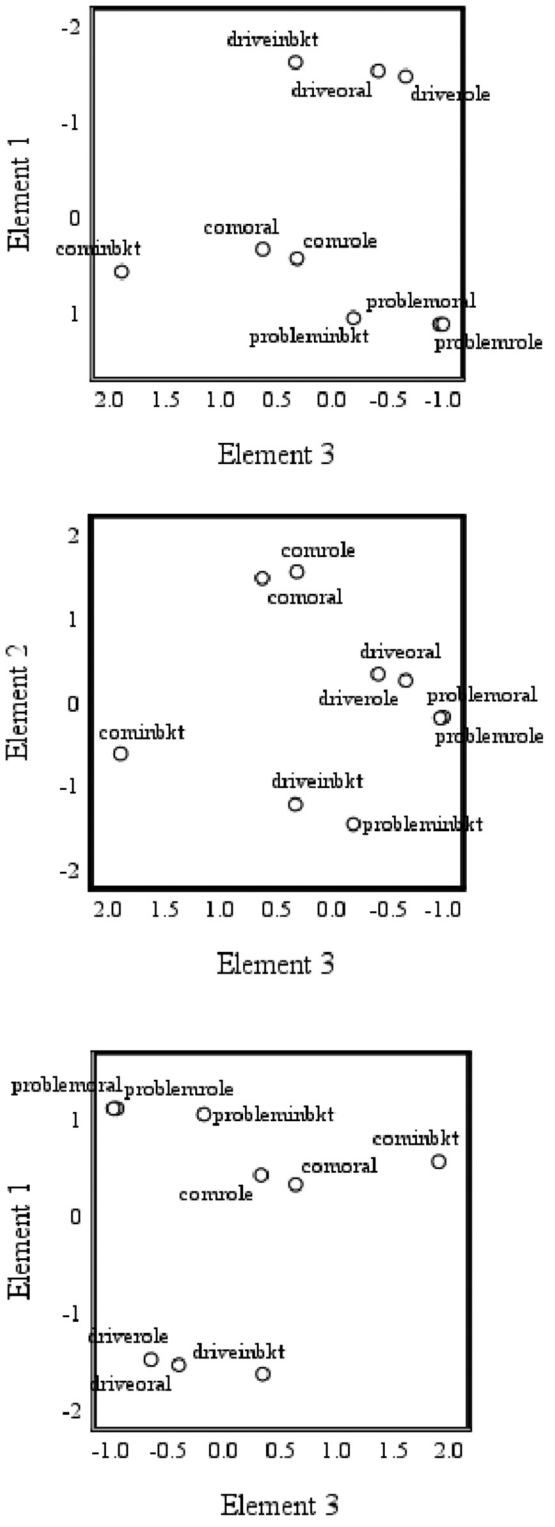

Figure 3. Three-element multidimensional scaling solution for assessment center dimension-exercise combinations. *Note*. The 3-element solution is shown above from different angles to facilitate interpretation. com = communication skills, problem = problem solving, oral = oral presentation, role = role play, inbkt = in-basket. Element 1 represents the expectation that dimensions should be stable across different exercises. Elements 2 and 3 differentiate different aspects of written from verbal behavioral output. Specifically, Element 2 differentiates communication in the role play and oral presentation from written communication in the in-basket. Element 3 primarily distinguishes communication in the in-basket from problem solving in the oral presentation and role play. Stress = .09, *R* square = .94.

Exercise-dimension combinations

The 3-element solution shown in Figure 3 reflects results for exercise-dimension combinations (see RQ3). Figure 3 suggests an outcome that incorporates an interpretation of both a comparative evaluation for dimensions and a continuum for exercises. Specifically, Element 1 suggests a clustering of same AC dimensions across different exercises. This is suggestive of the expectation that same AC dimension observations should be rated such that they coalesce across exercises (e.g., Handyside & Duncan, 1954; Lance et al., 2009). Elements 2 and 3 primarily distinguish written from verbal forms of behavioral output, which is suggestive of the media richness element observed in the substructure for exercises.

Discussion

The findings of our study, based on the results of an MDS analysis, contribute to the literature on AC dimension and exercise taxonomies from a largely unrepresented SME assessor perspective. On dimensions, our results, which we describe in detail below, suggest additions and refinements to existing taxonomies. On exercises, our results suggest a novel approach towards categorizing different exercises according to varying degrees of media richness and interpersonal interaction. Our findings generally offer new insights into theoretical bases for AC scores based on dimensions and exercises. We provide updated guidance to practitioners on how to differentiate between different dimensions and exercises for further research and testing in operational ACs. We furthermore demonstrate the application of a unique MDS methodological approach to modeling AC expert assessor cognitive representations of the dimensionality of AC dimensions and exercises.

The goals of our study involved applying MDS-derived cognitive representations to contribute to the development of generalized taxonomies for dimensions (RQ1), exercises (RQ2), and to advance an understanding of the application of dimension taxonomies by evaluating assessor expectations when dimensions and exercises are combined (RQ3). MDS offers unique suggestions about whether assessors perceive comparative evaluations (i.e., simple, distinct categories), continua (i.e., stimuli that differ by varying degrees), or a mixture of both perspectives (Hair et al., 1998). We found evidence for comparative evaluations for dimensions, continua for exercises, and both comparative evaluations and continua for dimension-exercise combinations, as we describe below.

As a brief overview of our results, we found novel evidence for a refinement of existing AC taxonomies focused on construct differentiation. In Table 7, we present our summary taxonomies for dimensions (RQ1) and exercises (RQ2), both of which present unique contributions to the AC literature. Regarding dimension-exercise combinations (RQ3), we found that assessors typically expected observations relevant to the same dimensions to coalesce across different exercises. We expand on and discuss these findings below in relation to previous research on dimension and exercise taxonomies.

Dimensions

Building on previously developed taxonomies for dimensions (Arthur et al., 2003; Meriac et al., 2014), our SME assessor perspective suggested a perceptual separation between five dimensions reflected in (a) task orientation, (b) interpersonal relations, (c) activity, (d) organizing and planning, and (e) tolerance for stress/uncertainty (RQ1, see Table 7). This conclusion was based on results from a 3-element[7] MDS solution for dimensions, involving the following distinctions in element 1: task orientation versus interpersonal relations; in element 2: organizing and planning versus problem

[7]Note that unless the analysis refers to continua, the number of elements does not necessarily equate to the number of summary clusters identified in the final solution. This is because elements could indicate distinctions between different stimuli (i.e., two stimuli per element) that appear more than once.

Table 7. Dimension and exercise taxonomies based on Assessor perspectives.

Score Type/Variable Title	Definition
Dimensions (simple separation)	
Task-orientation	Coordination and determination applied to work tasks.
Interpersonal relations	Consideration and empathic influence in interactions with other people.
Activity	Energy level maintained in response to work requirements.
Organizing and planning	Planning and allocation of resources applied to work activities.
Tolerance for stress/uncertainty	Maintaining performance on work tasks under pressure.
Exercises (continua)	
Media richness	Extent to which tasks reflect lean (e.g., written only) or rich content (e.g., containing visuo-auditory stimuli)
Interpersonal interaction	Extent to which exercises reflect low interdependence between participants (e.g., in a written exercise) or high interdependence (e.g., in a group exercise).

Definitions in this table were adapted from Thornton and Byham (1982), Meriac et al. (2014), Arthur et al. (2003), Thornton (1992), Hoffman et al. (2015), Dierdorff et al. (2012), and Maity et al. (2018).

solving, drive, and tolerance for stress/uncertainty; and in element 3: tolerance for stress/uncertainty versus drive (or activity).

In element 3, drive was separated from other dimensions. Akin to the findings of Arthur et al. (2003), there were subcomponents involved in our definition of drive (see Table 2). One aspect of this definition concerns performance standards and organizational advancement. We suggest that this aspect of drive is relevant to task orientation and is therefore likely relevant to element 1 (which includes reference to task orientation). The other aspect of our definition of drive is likely concerned with activity level. Given that general activity is not specific to task orientation, we suggest that the activity component of drive is the basis for further separation in element 3 (referred to as "energy" in Thornton & Byham, 1982, p. 139). For this reason, we refer to drive in element 3 as activity.

Our analyses suggest that several of the Arthur et al. (2003) dimensions were perceived as interrelated by our SME participants. This includes relationships observed between influencing others, consideration and awareness of others, communication skills, and problem solving, which could be due to conceptual similarities (e.g., between influencing others and consideration and awareness of others). However, it is possible that dimensions such as problem solving and communication are already implied in other dimensions. Problem solving, as defined in Thornton and Byham (1982), could be considered as an analogue of GMA. GMA might be considered necessary for and therefore implied in the operation of task orientation. Likewise, communication skills are likely implied in various manifestations of interpersonal relations. The suggestion here is that it would be redundant to evaluate problem solving as separate from task orientation or to evaluate communication as separate from interpersonal relations because of their implied nature.

Bearing in mind the suggestion of implied dimensions, there are both differences and similarities between our taxonomy and that of Arthur et al. (2003). Arthur et al. summarized six dimensions. Their consideration/awareness of others, drive, and organizing and planning summary dimensions find conceptual matches with our interpersonal relations, drive, and organization and planning dimensions. Arthur et al. suggested the summary dimensions communication and influencing others. Our analysis suggests both these dimensions can be subsumed into interpersonal relations. Arthur et al. also suggested a separate problem-solving dimension, which our analysis suggests is implied in organization and planning and task orientation. Our results moreover suggest some degree of distinction from other dimensions concerning tolerance for stress/uncertainty (e.g., Oral & Karakurt, 2022; Rolin et al., 2022).

As with the comparison to Arthur et al. (2003), our analysis raises points of both difference and similarity when compared to the structure suggested in the Meriac et al. (2014) study. Meriac et al. found evidence for three summary dimensions, including administrative skills, relational skills, and drive. Their administrative skills dimension is analogous to our organizing and planning dimension

and to a component of our task orientation dimension. The Meriac et al. relational skills dimension finds an analogue in our interpersonal relations dimension and both structures include a consideration of the drive summary dimension (see above).

In terms of a theoretical basis for our dimension structure, akin to Meriac et al. (2014), our distinction between task-orientation and interpersonal relations suggests an analogous distinction between initiating structure (or getting ahead: goal attainment, task orientation) and consideration (or getting along: concern for others' welfare, interpersonal interactions, see R. Hogan & Kaiser, 2005; Judge et al., 2004). On the drive dimension, Meriac et al. (2014) specifically refer to its association with getting ahead (from J. Hogan & Holland, 2003), which is analogous to our task-orientation dimension. However, as discussed, we suggest that part of the drive definition could be interpreted as a separate construct concerning activity level (also see Thornton & Byham, 1982).

Our identification of organizing and planning as a distinct dimension relates to administrative concepts covered in Meriac et al. (2014). In ACs, it is commonly the case that assessees are required to assimilate and integrate information that will be used strategically in the completion of simulated tasks (Povah & Thornton, 2011). Thus, it is of no surprise that organizing and planning appears both in our structure and that of Arthur et al. (2003).

Our findings around implied dimensions are particularly pertinent for practice. For example, communication skills are already implied in the interpersonal relations concept. Likewise, problem solving is implied in the organizing and planning concept. Evaluating implied dimensions is likely, according to our results, to lead to conceptual redundancy, as has been observed in previous AC literature (Bowler & Woehr, 2006). Our results for dimensions are oriented toward maximizing conditions for conceptual distinctiveness for practice and research. Nevertheless, we caution that there is no guarantee that applying the dimensions in Table 7 will result in reliable dimension scores. Because a within-exercise structure is routinely observed in AC ratings (Jackson et al., 2022), exercise scores may present a viable, although under-researched, alternative to dimension scores.

Exercises

Previous studies have directly applied the taxonomy of common AC exercises based on Thornton (1992) to classify different exercises from AC research (e.g., Hoffman et al., 2015; Lievens et al., 2006). Our aims extended this work by seeking evidence for a summary taxonomy for AC exercises as operationalized in the Thornton taxonomy from the perspective of experienced assessor SMEs (RQ2). Our results suggest a unique approach to classifying exercises based on a 2-element structure involving the summary continua (a) media richness and (b) interpersonal interaction.

Media richness, as suggested in our element 1 for exercises, has been discussed and developed in the literature on consumer behavior, suggesting a cross-disciplinary contribution to the AC literature. Maity et al. (2018) describe rich media as that which reflects visual or audible components. In the context of AC exercises, our analysis suggests this potentially includes exercises such as LGDs, role plays, and oral presentations. Rich media could also refer to video-based exercises or exercises that include recorded audio. In contrast to rich media, Maity et al. describe lean media as that which reflects sparse forms of stimuli, such as written text. Depending on their design and content, our analysis suggests this potentially includes written case analyses or in-baskets. It is possible that some exercises may include a combination of rich and lean elements. However, we suggest that an important aim when developing exercises is to retain job-relevant media (Thornton & Mueller-Hanson, 2004; Wernimont & Campbell, 1968).

Media richness finds some comparisons with the exercise characteristics component of the Hoffman et al. (2015) study. Hoffman et al. identified task complexity, structure, and fidelity (i.e., job-relevance) as being relevant to the classification of exercise characteristics, all of which could possibly relate, at least in part, to level of media richness. For example, exercises with higher fidelity may involve stimuli with a greater degree of richness and could include more detailed visual and auditory stimuli.

Element 2 in our analysis of exercises suggested a distinction by level of interpersonal interaction. This includes a differentiation to varying degrees between what are primarily individual or one-on-one exercises (role plays and oral presentations) from exercises that involve group interactions (LGDs and group-based case analyses). A topic relevant to these findings was raised in Dierdorff et al. (2012), where role expectations were described as differing by level of interdependence between employees. Hoffman et al. (2015) also identified AC exercise characteristics relevant to interpersonal interaction, including interdependence of participants and interpersonal orientation. Both these characteristics summarize levels of social interaction. Thus, our taxonomy assists in simplifying this component of the Hoffman et al. description of AC exercises.

Interpersonal interaction has been related to key components of role behavior in the research literature (Dierdorff et al., 2012). Thus, our findings support the idea that exercises can be developed as simulations of role behavior, as is discussed in the AC literature (Jackson, 2012; Lance, 2008). For example, a role play could be developed into an exercise oriented toward a supportive managerial role. We encourage further research into scoring ACs in this manner with each exercise representing a simulated work role that varies by degree of interpersonal interaction. On that note, AC architects could be guided by our findings on how to meaningfully distinguish between exercises. A key aim in ACs applied in employee selection is to predict job performance ratings (Hermelin et al., 2007). Redundancy in prediction could perhaps be minimized by selecting exercises that differ according to level of media richness and interpersonal interaction, depending on the requirements of the focal job (also see Speer, Christiansen, Goffin, et al., 2014).

We emphasize that that unlike our analysis involving dimensions, where we identified comparative evaluations, we found evidence for continua relating to our MDS solution for exercises. Specifically, levels of media richness and interpersonal interaction could vary by degree, depending on exercise design and content.

For practice, our results for AC exercises (as summarized in Table 7) provide a framework that can be used to support decisions about the inclusion of exercises in operational ACs. This guidance is relevant to both traditional dimension-based and task-based ACs, where, in the latter, exercise scores are applied (e.g., an overall score for a group discussion) instead of dimension scores (e.g., Lowry, 1997). Bearing in mind guidance based on job analysis and accepted AC practice guidelines (e.g., International Taskforce on Assessment Center Guidelines, 2015), our framework suggests that to assist in optimizing conceptual distinctiveness, practitioners may consider developing or including a mix of exercises that differ by degree of media richness and interpersonal interaction. For example, a practitioner may consider including, among other exercises, case analyses (if developed with low media richness and high interpersonal interaction) and oral presentations (if, in contrast, developed with high media richness and low interpersonal interaction). The aim here would be to cover as broadly as possible the criterion job performance domain so as to minimize redundancy and maximize the possibility of criterion-related validity (in keeping with the findings of Speer, Christiansen, Goffin, et al., 2014).

Dimension-exercise combinations and task-based perspectives

Regarding the evaluation of dimensions across exercises (RQ3), the results of element 1 in our MDS analysis (see Figure 3) suggested a clustering of same dimensions across different exercises. This suggests that assessors in our study were generally aware of formal scoring expectations in a dimension-based AC. These expectations possibly arise as a function of training oriented toward dimensions construed as constructs that "should" manifest in a relatively stable manner across different exercises (e.g., International Taskforce on Assessment Center Guidelines, 2015). However, results from many or most studies of AC measurement structure suggest that different dimensions tend to correlate strongly within exercises, and relatively weak correlations are often observed for same dimensions across exercises (Lance, 2008; Sackett & Dreher, 1982).

Our results thus suggest a disparity between what is expected by assessors (i.e., cross-exercise stable dimensions, see RQ3) versus what is regularly found to represent the structure of operational AC ratings (i.e., exercise factors, Jackson et al., 2022). This disparity suggests that, as a likely consequence of training, assessors develop schema about dimensions as cross-exercise-stable constructs that do not match operational AC ratings. These outcomes are striking because they imply that exercise factors tend to arise in ACs despite assessors developing, via formal training, schema around an entirely different, dimension-based scoring approach.

It has been suggested or implied in previous literature that clearer differentiations between AC dimensions may lead to improvements in their psychometric structure (e.g., Arthur et al., 2003; Meriac et al., 2014). Following this logic, Heimann et al. (2022) replaced dimensions with BFM personality traits in ACs and observed improved psychometric characteristics for their scoring approach. However, despite the common emergence of exercise factors in traditional dimension-based approaches, ACs are commonly found to predict work outcomes (e.g., Hermelin et al., 2007). Thus, ACs likely measure constructs of value, even if they are not those intended for evaluation. As a consequence, perhaps an alternative perspective on the internal measurement properties of ACs is required rather than a "fix" for ACs that do not measure what was intended by AC architects.

Following this logic, rather than attempting to "repair" ACs, another approach could be to develop an understanding about what they measure and formalize scoring around that structure. Several authors have suggested that an approach toward scoring performance constructs for each exercise that reflects the commonly observed structure of AC ratings (e.g., Lance et al., 2000; Lievens & Christiansen, 2012; Sackett & Dreher, 1982). This is often referred to as the task-based AC perspective (Jackson et al., 2005), where scoring occurs within, rather than across, exercises.

A criticism of the task-based approach is that it abandons a focus on substantive, psychological constructs in favor of a focus on assessee responses to non-substantive methods (Arthur & Villado, 2008). As a counter to this critique, researchers contributing to the task-based AC literature suggest that psychological mechanisms are likely to underlie the ability of ACs to predict performance (Jackson et al., 2010). However, those psychological mechanisms are, in the task-based view, unlikely to be those formalized for scoring in ACs in the form of dimensions. Rather, they are possibly a consequence of personality, cognitive ability, and/or other psychological characteristics expressed in exercise performance (Jackson et al., 2016). In support of this idea, several researchers have found relationships between overall AC ratings and external measures of personality and cognitive ability (e.g., Furnham et al., 2008; Schmidt & Hunter, 1998).

It is our hope that the taxonomy developed in the present study for exercises can assist in the development of a theoretical basis for task-based ACs and, by extension, ACs generally. It is moreover hoped that our exercise taxonomy can assist researchers and practitioners adopting a task-based approach to include exercises in their ACs that are differentiated on the elements we have identified. This could potentially assist in reducing the potential for redundancy among exercise scores, thereby maximizing the potential to predict work-related outcomes with ACs.

Limitations and future directions

Our data were gleaned wholly from SME assessors, building on previous research where the level of assessor experience was uncertain (e.g., Arthur et al., 2003; Hoffman et al., 2015; Meriac et al., 2014). Our observations were not based on aggregated AC ratings and were therefore not afflicted by conflated variance sources, as is relevant to some previous work in this area (e.g., Lenzenweger, 2015; The OSS Assessment Staff, 1948). Our analyses were moreover not reliant on the structural soundness of AC ratings in previous studies. However, we were not able to ascertain whether our results are likely to make a difference to structures in ratings generated in operational ACs. We suggest that a key problem with AC rating structures is that insufficient a priori developmental work has been produced relating to

the AC measurement design in advance of AC participants being rated in ACs. It is the literature on the development of a priori taxonomies to which we offer a contribution. We hope that, in operational ACs, future researchers will formally test and further refine the taxonomies suggested by our MDS studies.

On the point of including expert assessors as participants, one consideration from the expert-novice literature is that experts can potentially become cognitively entrenched in their domains of interest (Phan & Ngu, 2021). A possible consequence of cognitive entrenchment is a lack of flexibility and adaptability (Dane, 2010). We are unaware of any studies that have explored cognitive entrenchment as it relates to ACs, and this could present a fruitful area for future research. Of some reassurance pertaining to the present study, Dane posits that entrenchment can potentially be alleviated where experts are "engaging in a dynamic environment and attending to outside-domain tasks" (pp. 594–-595). Given that assessing in an AC is unlikely to constitute the sole task of those serving that role (AC assessors are also often managers, Krause et al., 2011; Lowry, 1996), assessors may be routinely exposed to a broad range of work activities. Furthermore, ACs are often found to predict performance (Sackett et al., 2022). Thus, despite any possible cognitive entrenchment among assessors, ACs still likely offer information of value.

A key aim in our study was to provide background development that could potentially lead to an increase in the proportion of variance explained in the true (or universe) score component of dimension or exercise scores in future AC research. We emphasize, however, that increasing the proportion true score does not equate to substantially or even detectably increasing the reliability of scores generated by an AC. Regardless, we argue that efforts directed toward psychometric improvements, particularly in dimension-based evaluation, are needed in I-O psychology if researchers and practitioners are to continue applying dimension and competency scores (Dewberry, 2024; Jackson et al., 2022; Lance, 2008).

Our analyses relating to dimensions and exercises included only a subset of dimensions and exercises from the AC literature. Moreover, an even smaller subset of dimensions and exercises were included in our analysis of dimension-exercise combinations. This raises questions about the potential for bias arising from the inclusion of a specific set of dimensions and exercises in our MDS tasks. However, as a basis for our MDS tasks, we drew on dimension stimuli from Arthur et al. (2003). In the Arthur et al. study, 168 dimension titles were coded from the literature. This original basis was not obviously restricted. We drew on exercise stimuli from Hoffman et al. (2015), which was based on the taxonomy from Thornton (1992). The Thornton taxonomy was, itself, based on a practice survey from Gaugler et al. (1990) and reflects exercises similar to those presented in other practice surveys (e.g., Eurich et al., 2009, Krause et al., 2011; Spychalski et al., 1997). Again, as a basis, this is not obviously restricted. Thus, our take is that effective groundwork on a priori taxonomies has already been achieved in previous research and our study extends that work by focusing on furthering construct differentiation via expert assessor SME perspectives.

On a related note, regarding dimension-exercise combinations, our aim was to offer a demonstration of assessor expectations about an illustrative and relevant set of dimensions and exercises. This aim does not imply a desire to generalize our dimension-exercise combinations to a universe of all such combinations.

Cross-taxonomy comparisons and conclusions

We aimed to contribute to the development of dimension and exercise taxonomies from the SME assessor perspective via MDS methodology (see Table 7 for a summary of our taxonomies along with definitions). Our proposed refinements reveal similarities and differences when comparing across the extant taxonomies summarized in Table 8. Regarding the number of summary dimensions, our taxonomy (with five dimensions) presents something of a mid-way between those of Arthur et al. (2003, six to seven dimensions) and Meriac et al. (2014, three dimensions). Of note in Table 8 is that, congruent with Meriac et al., our interpersonal relations dimension subsumes three dimensions from

the Arthur et al. framework. However, our task-orientation and organizing and planning dimensions offer a more specific structure than that suggested in the broad administrative skills dimension in Meriac et al. (2014).

Regarding exercise taxonomies (also shown in Table 8), our framework included the same exercise categories as those in Hoffman et al. (2015). However, our taxonomy suggests an entirely unique approach toward considering exercises based on their content and design. This perspective suggests two elements that vary by degree: level of media richness and level of interpersonal interaction. Responses in the current study were based on specific exercise descriptions provided to respondents. However, we anticipate that exercise content and design will differ depending on the aims of the AC architect and the organization adopting the evaluation approach. To illustrate, given our exercise definitions, we found results suggesting that case analyses were low on media richness, but high on interpersonal interaction. However, case analyses could very reasonably be designed to include rich media and opportunities for interpersonal interaction, if that suited organizational aims. The main point raised in our taxonomy is that levels of media richness and interpersonal interaction possibly represent a key design consideration, regardless of the exercise category concerned.

We suggest that our results refine areas of overlap in previous work in this area and offer a theoretical and practical direction for dimension and exercise scoring perspectives. This includes background knowledge that could apply to the development of dimension- and exercise-based ACs. We moreover found evidence that assessors in our study had developed a schema around the evaluation of same dimensions across different exercises. This is despite the finding that operational AC ratings rarely reflect a pattern that matches this schema. We hope our AC taxonomies, based on expert assessor SME judgments, will be applied, tested, and further developed in future research on the AC method.

Table 8. Comparison of assessment center dimension and exercise taxonomies.

Dimensions	Arthur et al. (2003)	Meriac et al. (2014)	Current study
	Problem solving	Administrative skills	Task-orientation
	Organizing and planning		Organizing and planning
	Drive	Drive	Activity
	Communication	Relational skills	Interpersonal relations
	Consideration/awareness of others		
	Influencing others		
	(Tolerance for stress/uncertainty)	NA	Tolerance for stress/uncertainty
Exercises	**Thornton (1992)**	**Hoffman et al. (2015)**	**Current study**
	In-baskets	In-baskets	In-baskets, low MR, low II
	Group tasks	LGDs	LGDs, high MR, high II
	Group discussions (assigned positions)		
	Group discussions (unassigned positions)		
	Interview simulations[a]	Role plays	Role plays, high MR, low II
	Business games	Case analyses	Case analyses, low MR, high II
	Case analyses		
	Fact finding exercises		
	Oral presentations	Oral presentations	Oral presentations high MR, low II

MR = media richness, II = interpersonal interaction, LGD = leaderless group discussion. NA = not applicable. We suggest that levels of MR and II depend on specific exercise content. Tolerance for stress/uncertainty appears in parentheses because this dimension was not included in the final Arthur et al. (2003) taxonomy. It is not implied above that the Arthur et al. dimension taxonomy maps neatly on to the current dimension taxonomy due to differences between cross-study outcomes. For example, task orientation in the current study is not assumed to replace problem solving in Arthur et al. Rather, problem solving is interpreted as being implied in other dimensions. [a]Defined in Thornton (1992) as "adaptations of the role-playing technique" (p. 69).

Acknowledgments

This study was sponsored by the Douglas W. Bray and Ann Howard Research Grant via the SIOP (Society for Industrial and Organizational Psychology) Foundation. Portions of these findings were presented as part of a symposium at the 2022 SIOP conference, Seattle, Washington, USA.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was sponsored by the Douglas W. Bray and Ann Howard Research Grant via the SIOP (Society for Industrial and Organizational Psychology) Foundation.

Data availability statement

Research data are available on request.

References

Anderson, N., Lievens, F., van Dam, K., & Born, M. (2006). A construct-driven investigation of gender differences in a leadership-role assessment center. *Journal of Applied Psychology*, *91*(3), 555–566. https://doi.org/10.1037/0021-9010.91.3.555

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA publications and communications board task force report. *The American Psychologist*, *73*(1), 3–25. https://doi.org/10.1037/amp0000191

Arthur, W., Jr. (2012). Dimension-based assessment centers: Theoretical perspectives. In D. J. R. Jackson, C. E. Lance, & B. J. Hoffman (Eds.), *The psychology of assessment centers* (pp. 95–120). Routledge.

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*(1), 125–154. https://doi.org/10.1111/j.1744-6570.2003.tb00146.x

Arthur, W., Jr., & 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*(2), 435–442. https://doi.org/10.1037/0021-9010.93.2.435

Barrick, M. R., & Mount, M. K. (2005). Yes, personality matters: Moving on to more important matters. *Human Performance*, *18*(4), 359–372. https://doi.org/10.1207/s15327043hup1804_3

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, *9*(1/2), 9–30. https://doi.org/10.1111/1468-2389.00160

Blustein, D. L., Smith, C. M., Wu, X., Guarino, P. A., Joyner, E., Milo, L., & Bilodeau, D. C. (2022). "Like a tsunami coming in fast": A critical qualitative study of precarity and resistance during the pandemic. *Journal of Counseling Psychology*, *69*(5), 565–577. https://doi.org/10.1037/cou0000615

Borman, W. C. (2012). Dimensions, tasks, and mixed models: An analysis of three diverse perspectives on assessment centers. In D. J. R. Jackson, C. E. Lance, & B. J. Hoffman (Eds.), *The psychology of assessment centers* (pp. 309–320). Routledge.

Borman, W. C., & Brush, D. H. (1993). More progress toward a taxonomy of managerial performance requirements. *Human Performance*, *6*(1), 1–21. https://doi.org/10.1207/s15327043hup0601_1

Borman, W. C., & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. *Human Performance*, *10*(2), 99–109. https://doi.org/10.1207/s15327043hup1002_3

Bowler, M. C., & Woehr, D. J. (2006). A meta-analytic evaluation of the impact of dimension and exercise factors on assessment center ratings. *Journal of Applied Psychology*, *91*(5), 1114–1124. https://doi.org/10.1037/0021-9010.91.5.1114

Cook, M. (2016). *Personnel selection: Adding value through people* (6th ed.). Wiley Blackwell.

Dane, E. (2010). Reconsidering the trade-off between expertise and flexibility: A cognitive entrenchment perspective. *Academy of Management Review*, *35*(4), 579–603. https://doi.org/10.5465/AMR.2010.53502832

Dewberry, C. (2024). Assessment centers do not measure competencies: Why this is now beyond reasonable doubt. *Industrial and Organizational Psychology*, *17*(2), 154–175. https://doi.org/10.1017/iop.2024.5

Dewberry, C., & Jackson, D. J. R. (2016). The perceived nature and incidence of dysfunctional assessment center features and processes. *International Journal of Selection and Assessment*, *24*(2), 189–196. https://doi.org/10.1111/ijsa.12140

Dierdorff, E. C., Rubin, R. S., & Bachrach, D. G. (2012). Role expectations as antecedents of citizenship and the moderating effects of work context. *Journal of Management*, *38*(2), 573–598. https://doi.org/10.1177/0149206309359199

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

Eurich, T. L., Krause, D. E., Cigularov, K., & Thornton, G. C., III. (2009). Assessment centers: Current practices in the United States. *Journal of Business & Psychology*, *24*(4), 387–407. https://doi.org/10.1007/s10869-009-9123-3

Furnham, A., Jensen, T., & Crump, J. (2008). Personality, intelligence and assessment centre expert ratings. *International Journal of Selection and Assessment*, *16*(4), 356–365. https://doi.org/10.1111/j.1468-2389.2008.00441.x

Gatewood, R. D., Feild, H. S., & Barrick, M. R. (2016). *Human resource selection* (8th ed.) Cengage Learning.

Gaugler, B. B., Bentson, C., & Pohley, K. (1990). *A survey of assessment center practices in organizations* [Unpublished manuscript.

Goodge, P. (1988). Task-based assessment. *Journal of European Industrial Training*, *12*(6), 22–27. https://doi.org/10.1108/eb002258

Greenhaus, J. H., & Callanan, G. A. (Eds.) (2006). *Encyclopedia of career development* (Vol. 1). Sage Publications.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). Prentice Hall.

Handler, L. (2001). Assessment of men: Personality assessment goes to war by the office of strategic services assessment staff. *Journal of Personality Assessment*, *76*(3), 556–578. https://doi.org/10.1207/S15327752JPA7603_13

Handyside, J. D., & Duncan, D. C. (1954). Four years later: A follow-up of an experiment in selecting supervisors. *Occupational Psychology*, *28*, 9–23. https://psycnet.apa.org/record/1955-01641-001

Heimann, A. L., Ingold, P. V., Lievens, F., Melchers, K. G., Keen, G., & Kleinmann, M. (2022). Actions define a character: Assessment centers as behavior-focused personality measures. *Personnel Psychology*, *75*(3), 675–705. https://doi.org/10.1111/peps.12478

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*(4), 405–411. https://doi.org/10.1111/j.1468-2389.2007.00399.x

Hoffman, B. J., Kennedy, C. L., LoPilato, A. C., Monahan, E. L., & Lance, C. E. (2015). A review of the content, criterion-related, and construct-related validity of assessment center exercises. *Journal of Applied Psychology*, *100*(4), 1143–1168. https://doi.org/10.1037/a0038707

Hogan, J., & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. *Journal of Applied Psychology*, *88*(1), 100–112. https://doi.org/10.1037/0021-9010.88.1.100

Hogan, R., & Kaiser, R. B. (2005). What we know about leadership. *Review of General Psychology*, *9*(2), 169–180. https://doi.org/10.1037/1089-2680.9.2.169

IBM Corp. (2023). *IBM SPSS statistics for Windows. In (version 29)*.

International Taskforce on Assessment Center Guidelines. (2015). Guidelines and ethical considerations for assessment center operations. *Journal of Management*, *41*(4), 1244–1273. https://doi.org/10.1177/0149206314567780

Jackson, D. J. R. (2012). Task-based assessment centers: Theoretical perspectives. In D. J. R. Jackson, C. E. Lance, & B. J. Hoffman (Eds.), *The psychology of assessment centers* (pp. 173–189). Routledge/Taylor & Francis Group.

Jackson, D. J. R., Ahmad, M. H., Grace, G. M., & Yoon, J. (2011). An alternative take on assessment center research and practice: Task-based assessment centers. In N. Povah & G. C. Thornton III (Eds.), *Assessment centres and global talent management* (pp. 33–46). Gower Publishing.

Jackson, D. J. R., Michaelides, G., Dewberry, C., Nelson, J., & Stephens, C. (2022). Reliability in assessment centres depends on general and exercise performance, but not on dimensions. *Journal of Occupational & Organizational Psychology*, *95*(4), 739–757. https://doi.org/10.1111/joop.12398

Jackson, D. J. R., Michaelides, M., Dewberry, C., & Kim, Y. (2016). Everything that you have ever been told about assessment center ratings is confounded. *Journal of Applied Psychology*, *101*(7), 976–994. https://doi.org/10.1037/apl0000102

Jackson, D. J. R., Stillman, J. A., & Atkins, S. G. (2005). Rating tasks versus dimensions in assessment centers: A psychometric comparison. *Human Performance*, *18*(3), 213–241. https://doi.org/10.1207/s15327043hup1803_2

Jackson, D. J. R., Stillman, J. A., & Englert, P. (2010). Task-based assessment centers: Empirical support for a systems model. *International Journal of Selection and Assessment*, *18*(2), 141–154. https://doi.org/10.1111/j.1468-2389.2010.00496.x

Judge, T. A., Piccolo, R. F., & Ilies, R. (2004). The forgotten ones? The validity of consideration and initiating structure in leadership research. *Journal of Applied Psychology*, *89*(1), 36–51. https://doi.org/10.1037/0021-9010.89.1.36

Kleinmann, M., & Ingold, P. V. (2019). Toward a better understanding of assessment centers: A conceptual review. *Annual Review of Organizational Psychology & Organizational Behavior*, *6*(1), 349–372. https://doi.org/10.1146/annurev-orgpsych-012218-014955

Knezek, G., Gibson, D., Christensen, R., Trevisan, O., & Carter, M. (2023). Assessing approaches to learning with nonparametric multidimensional scaling. *British Journal of Educational Technology*, *54*(1), 126–141. https://doi.org/10.1111/bjet.13275

Kolk, N. J., Born, M. P., & van der Flier, H. (2004). A triadic approach to the construct validity of the assessment center - the effect of categorizing dimensions into a feeling, thinking, and power taxonomy. *European Journal of Psychological Assessment*, *20*(3), 149–156. https://doi.org/10.1027/1015-5759.20.3.149

Krause, D. E., Rossberger, R. J., Dowdeswell, K., Venter, N., & Joubert, T. (2011). Assessment center practices in South Africa. *International Journal of Selection and Assessment*, *19*(3), 262–275. https://doi.org/10.1111/j.1468-2389.2011.00555.x

Krause, D. E., & Thornton, G. C., III. (2009). A cross-cultural look at assessment center practices: Survey results from Western Europe and North America. *Applied Psychology: An International Review*, *58*(4), 557–585. https://doi.org/10.1111/j.1464-0597.2008.00371.x

Laher, S. (2013). Understanding the five-Factor Model and five-Factor theory through a South African cultural lens. *South African Journal of Psychology*, *43*(2), 208–221. https://doi.org/10.1177/0081246313483522

Lance, C. E. (2008). Why assessment centers do not work the way they are supposed to. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *1*(1), 84–97. https://doi.org/10.1111/j.1754-9434.2007.00017.x

Lance, C. E. (2012). Research into task-based assessment centers. In D. J. R. Jackson, C. E. Lance, & B. J. Hoffman (Eds.), *The psychology of assessment centers* (pp. 218–233). Routledge/Taylor & Francis Group.

Lance, C. E., Baranik, L. E., Lau, A. R., & Scharlau, E. A. (2009). If it ain't trait it must be method: (Mis)application of the multitrait-multimethod design in organizational research. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp. 337–360). Routledge/Taylor & Francis Group.

Lance, C. E., Foster, M. R., Nemeth, Y. M., Gentry, W. A., & Drollinger, S. (2007). Extending the nomological network of assessment center construct validity: Prediction of cross-situationally consistent and specific aspects of assessment center performance. *Human Performance*, *20*(4), 345–362. https://doi.org/10.1080/08959280701522031

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*(4), 323–353. https://doi.org/10.1207/S15327043HUP1304_1

Lenzenweger, M. F. (2015). Factors underlying the psychological and behavioral characteristics of office of strategic services candidates: The assessment of men data revisited. *Journal of Personality Assessment*, *97*(1), 100–110. https://doi.org/10.1080/00223891.2014.935980

Lievens, F. (2001). Assessors and use of assessment centre dimensions: A fresh look at a troubling issue. *Journal of Organizational Behavior*, *22*(3), 203–221. https://doi.org/10.1002/job.65

Lievens, F. (2008). What does exercise-based assessment really mean? *Industrial and Organizational Psychology-Perspectives on Science and Practice*, *1*(1), 112–115. https://doi.org/10.1111/j.1754-9434.2007.00020.x

Lievens, F. (2009). Assessment centres: A tale about dimensions, exercises, and dancing bears. *European Journal of Work & Organizational Psychology*, *18*(1), 102–121. https://doi.org/10.1080/13594320802058997

Lievens, F., Chasteen, C. S., Day, E. A., & Christiansen, N. D. (2006). Large-scale investigation of the role of trait activation theory for understanding assessment center convergent and discriminant validity. *Journal of Applied Psychology*, *91*(2), 247–258. https://doi.org/10.1037/0021-9010.91.2.247

Lievens, F., & Christiansen, N. D. (2012). Core debates in assessment center research: Dimensions 'versus' exercises. In D. J. R. Jackson, C. E. Lance, & B. J. Hoffman (Eds.), *The psychology of assessment centers* (pp. 68–91). Routledge.

Lowry, P. E. (1996). A survey of the assessment center process in the public sector. *Public Personnel Management*, *25*(3), 307–321. https://doi.org/10.1177/009102609602500304

Lowry, P. E. (1997). The assessment center process: New directions. *Journal of Social Behavior & Personality*, *12*(5), 53–62.

Maity, M., Dass, M., & Kumar, P. (2018). The impact of media richness on consumer information search and choice. *Journal of Business Research*, *87*(2), 36–45. https://doi.org/10.1016/j.jbusres.2018.02.003

Meriac, J. P., Hoffman, B. J., & Woehr, D. J. (2014). A conceptual and empirical review of the structure of assessment center dimensions. *Journal of Management*, *40*(5), 1269–1296. https://doi.org/10.1177/0149206314522299

Meriac, J. P., Hoffman, B. J., Woehr, D. J., & Fleisher, M. S. (2008). Further evidence for the validity of assessment center dimensions: A meta-analysis of the incremental criterion-related validity of dimension ratings. *Journal of Applied Psychology*, *93*(5), 1042–1052. https://doi.org/10.1037/0021-9010.93.5.1042

Merkulova, N., Melchers, K. G., Kleinmann, M., Annen, H., & Tresch, T. S. (2016). A test of the generalizability of a recently suggested conceptual model for assessment center ratings. *Human Performance*, *29*(3), 226–250. https://doi.org/10.1080/08959285.2016.1160093

Meyers, L. S., Gamst, G. C., & Guarino, A. J. (2013). *Performing data analysis using IBM SPSS*. Wiley.

Moss, S. E., Valenzi, E. R., & Taggart, W. (2003). Are you hiding from your boss? The development of a taxonomy and instrument to assess the feedback management behaviors of good and bad performers. *Journal of Management*, *29*(4), 487–510. https://doi.org/10.1016/S0149-2063%2803%2900022-9

Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *The American Psychologist*, *51*(2), 77–101. https://doi.org/10.1037/0003-066X.51.2.77

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.) McGraw-Hill.

Oral, M., & Karakurt, N. (2022). The impact of psychological hardiness on intolerance of uncertainty in university students during the COVID-19 pandemic. *Journal of Community Psychology*, *50*(8), 3574–3589. https://doi.org/10.1002/jcop.22856

The OSS Assessment Staff. (1948). *Assessment of men: Selection of personnel for the office of strategic services*. Rinehart & Company Inc.

Phan, H., & Ngu, B. (2021). A case for cognitive entrenchment: To achieve optimal best, taking into account the importance of perceived optimal efficiency and cognitive load imposition. *Frontiers in Psychology*, *12*, 1–19. https://doi.org/10.3389/fpsyg.2021.662898

Povah, N., & Thornton, G. C., III (Eds.) (2011). *Assessment centres and global talent management*. Gower.

Pulakos, E. D., Arad, S., Donovan, M. A., & Plamondon, K. E. (2000). Adaptability in the workplace: Development of a taxonomy of adaptive performance. *Journal of Applied Psychology*, *85*(4), 612–624. https://doi.org/10.1037/0021-9010.85.4.612

Putka, D. J., & Hoffman, B. J. (2013). Clarifying the contribution of assessee-, dimension-, exercise-, and assessor-related effects to reliable and unreliable variance in assessment center ratings. *Journal of Applied Psychology*, *98*(1), 114–133. https://doi.org/10.1037/a0030887

Roch, S. G., & O'Sullivan, B. J. (2003). Frame of reference rater training issues: Recall, time and behavior observation training. *International Journal of Training and Development*, *7*(2), 93–107. https://doi.org/10.1111/1468-2419.00174

Rolin, S. N., Flis, A., & Davis, J. J. (2022). Work coping, stress appraisal, and psychological resilience: Reaction to the COVID-19 pandemic among health care providers. *Psychology & neuroscience*, *15*(2), 131–146. https://doi.org/10.1037/pne0000257

Sackett, P. R., & Dreher, G. F. (1982). Constructs and assessment center dimensions: Some troubling empirical findings. *Journal of Applied Psychology*, *67*(4), 401–410. https://doi.org/10.1037/0021-9010.67.4.401

Sackett, P. R., Shewach, O. R., & Keiser, H. N. (2017). Assessment centers versus cognitive ability tests: Challenging the conventional wisdom on criterion-related validity. *Journal of Applied Psychology*, *102*(10), 1435–1447. https://doi.org/10.1037/apl0000236

Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range [personnel management & selection & training 3620]. *Journal of Applied Psychology*, *107*(11), 2040–2068. https://doi.org/10.1037/apl0000994

Schippmann, J. S., Ash, R. A., Battista, M., Carr, L., Eyde, L. D., Hesketh, B., & Sanchez, J. I. (2000). The practice of competency modeling. *Personnel Psychology*, *53*(3), 703–740. https://doi.org/10.1111/j.1744-6570.2000.tb00220.x

Schmidt, F. L., & Hunter, J. E. (1993). Tacit knowledge, practical intelligence, general mental ability, and job knowledge. *Current Directions in Psychological Science*, *2*(1), 8–9. https://doi.org/10.1111/1467-8721.ep10770456

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*(2), 262–274. https://doi.org/10.1037/0033-2909.124.2.262

Speer, A. B., Christiansen, N. D., Goffin, R. D., & Goff, M. (2014). Situational bandwidth and the criterion-related validity of assessment center ratings: Is cross-exercise convergence always desirable? *Journal of Applied Psychology*, *99*(2), 282–295. https://doi.org/10.1037/a0035213

Speer, A. B., Christiansen, N. D., Melchers, K. G., König, C. J., & Kleinmann, M. (2014). Establishing the cross-situational convergence of the ability to identify criteria: Consistency and prediction across similar and dissimilar assessment center exercises. *Human Performance*, *27*(1), 44–60. https://doi.org/10.1080/08959285.2013.854364

Spicer, J. (2005). *Making sense of multivariate data analysis*. Sage.

Spychalski, A. C., Quinones, M. A., Gaugler, B. B., & Pohley, K. (1997). A survey of assessment center practices in organizations in the United States. *Personnel Psychology*, *50*(1), 71–90. https://doi.org/10.1111/j.1744-6570.1997.tb00901.x

Stogdill, R. M. (1950). Leadership, membership and organization. *Psychological Bulletin*, *47*(1), 1–14. https://doi.org/10.1037/h0053857

Takane, Y., Young, F. W., & de Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika*, *42*(1), 7–67. https://doi.org/10.1007/BF02293745

Thoresen, C. J., & Thoresen, J. D. (2012). How to design and implement a task-based assessment center. In D. J. R. Jackson, C. E. Lance, & B. J. Hoffman (Eds.), *The psychology of assessment centers* (pp. 190–217). Routledge.

Thornton, G. C., III. (1992). *Assessment centers in human resource management*. Addison-Wesley.

Thornton, G. C., III, & Byham, W. C. (1982). *Assessment centers and managerial performance*. Academic Press.

Thornton, G. C., III, & Mueller-Hanson, R. A. (2004). *Developing organizational simulations: A guide for practitioners and students*. Routledge.

Tross, S. A., & Maurer, T. J. (2000). The relationship between SME job experience and job analysis ratings: Findings with and without statistical control. *Journal of Business & Psychology*, *15*(1), 97–110. https://doi.org/10.1023/A:1007770919305

Tziner, A., Fein, E. C., Kim, S.-K., Vasiliu, C., & Shkoler, O. (2020). Combining associations between emotional intelligence, work motivation, and organizational justice with counterproductive work behavior: A profile analysis via multidimensional scaling (PAMS) approach. *Frontiers in Psychology*, *11*, 1–18. https://doi.org/10.3389/fpsyg.2020.00851

Vera, J. F., & Macias, R. (2021). On the behaviour of K-means clustering of a dissimilarity matrix by means of full multidimensional scaling. *Psychometrika*, *86*(2), 489–513. https://doi.org/10.1007/s11336-021-09757-2

Wernimont, P. F., & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology*, *52*(5), 372–376. https://doi.org/10.1037/h0026244

Young, F. W., & Harris, D. F. (2012). Multidimensional scaling. In M. J. Norusis (Ed.), *SPSS statistics 19.0: Advanced statistical procedures companion* (pp. 361–430). Pearson Prentice Hall.