Psihologia Resurselor Umane, **22** (2024), 6–17

https://doi.org/10.24837/pru.v22i1.545

RESEARCH ARTICLE

Dimension- or Task-based Assessment Centers? A direct comparison study of two measurement approaches

IOANA BOȘNEAG*
Department of Psychology and Cognitive Science, University of Bucharest

DRAGOȘ ILIESCU
Department of Psychology and Cognitive Science, University of Bucharest
Department of Industrial Psychology, Stellenbosch University

Abstract

This study investigates two divergent strategies for assessment center ratings, dimension-based assessment centers (DBAC) and task-based assessment centers (TBAC), in order to directly compare their model fit using the same data. The sample consisted of 126 Romanian students who took part in an Assessment and Development Center. The assessment center matrix included two exercises, analysis and in-tray, and two traditional dimensions, problem solving and organizing and planning, as well as one task-based dimension. Findings confirmed that the TBAC model performs significantly better than the DBAC model. Implications regarding the construct validity of TBAC ratings and internal reliability of DBAC ratings are discussed.

Keywords

assessment centers, dimension-based, task-based, confirmatory factor analysis

Assessment Centers (ACs) have gained much popularity among both practitioners and researchers over the past several decades. They are highly valued in personnel selection for both what they measure and how they measure it (Lievens & Thornton, 2017; Thornton & Rupp, 2006). The behavioral dimensions assessed within ACs are directly linked to job performance criteria. These are measured in different exercises that simulate tasks related to the job. Following a multitrait-multimethod (MTMM) approach derived from a thorough job analysis (Campbell & Fiske, 1959), the ACs' outcomes are of high practical relevance in regard to both prediction of future performance and developmental feedback (Arthur et al., 2003; Gaugler et al., 1987; Sackett et al., 2017; Schmidt & Hunter, 2016).

Notwithstanding, a significant amount of research has revealed evidence that does not completely support a MTMM approach in ACs. The debate was started by Sackett and Dreher (1982) and was continued by other researchers (Arthur et al., 2008; Bowler &

* Correspondence concerning this article should be addressed to Ioana Boșneag, Department of Psychology and Cognitive Science, University of Bucharest, Sos. Panduri Nr. 90, 050657 Bucharest, Romania. E-mail: ioana.bosneag@drd.unibuc.ro

Woehr, 2006; Cahoon et al., 2012; Fleenor, 1996; Hoffman et al., 2011; Kuncel & Sackett, 2014; Lance, 2008) to conclude that AC dimension ratings are not consistent across all the exercises in which they are designed to be measured, but that instead these are rather situational, relative to a certain context. This conclusion questioned the construct validity of AC ratings, as well as their reliability and led to two broad schools of thought in regard to how these behavioral constructs should be operationalized. One of them is in favor of a traditional dimension-based approach towards behavioral ratings (DBAC) (Arthur et al., 2003; Bowler & Woehr, 2009; Meriac et al., 2008), while the other promotes a task-based approach (TBAC) focused on performance in exercises (Goodge, 1988; Hoffman et al., 2011; Jackson et al., 2010, 2016; Jackson & Englert, 2011).

While the TBAC approach seems to be an alternative to DBAC, the two models have so far been studied only independently, i.e., without being directly compared against each other. Our study addresses this gap by comparing the two divergent assessment strategies in ACs, and judging their relative merit based on model fit measured against the same data sample.

What do DBACs and TBACs measure?

DBACs have traditionally been used in organizational practice, and most psychometric characteristics of AC ratings have been studied through this model. Driven by the theoretical MTMM approach (Campbell & Fiske, 1959), DBACs are designed based on an AC matrix that contains two components: traits and methods, defined as dimensions and exercises. The AC matrix is derived from an analysis of the targeted job profile, which offers the relevant input for operationalizing dimensions as well as for the selection of tasks that are then to be simulated in exercises (Thornton & Rupp, 2006).

Each dimension aims at measuring a specific competency, described by a range of behaviors that are expected to lead to the desired performance, if they are efficiently manifested in tasks. Applying the MTMM principles, each dimension should be assessed in more than one exercise (Rupp et al., 2015). Thus, each behavior is measured within each exercise resulting in a post exercise dimension rating (PEDR). PEDRs are quantitatively and/or qualitatively integrated in an overall rating per dimension and, further, in an overall assessment rating (OAR) (Lievens et al., 2008; Kleinmann & Ingold, 2019).

Dimensions measure what a person is able to do in specific performance situations, and not what they *could* do or their potential to perform (Arthur et al., 2008). A wide range of aspects may be taken into consideration when defining dimensions, such as the capability to find solutions to problems, provide performance feedback to an employee, hold a business presentation etc. Meta-analytic studies suggested that the constructs for DBAC could be grouped in 6 narrow categories: (a) consideration/ awareness of others, (b) communication, (c) drive, (d) influencing others, (e) organizing and planning, and (f) problem solving, being emphasized a true criterion-related validity ranging from .25 to .39 (Arthur et al., 2003). In addition, a good criterion-related validity was highlighted for OARs, larger than and incremental over other psychological traits (Gaugler et. al., 1987; Meriac et. al., 2008; Sackett et. al., 2017; Schmidt & Hunter, 2016).

TBAC came out as an alternative to the traditional DBAC (Gorham, 1978; Jackson et al., 2010), promoting a 'dimensionless' design approach to AC (Iles, 1992). It attributes a greater importance to the effects of exercises, focusing on measuring merely the behaviors directly involved to perform the tasks. The design of TBAC relies on tasks resulted from job analysis, which are simulated in exercise and then operationalized into specific behaviors. Behaviors might differ from one exercise to another, and they do not necessary underlie a specific trait. They are rated within the exercise and, afterwards, integrated in an overall exercise rating and OAR computed based on exercise ratings (Jackson et al., 2010).

The similarities as well as differences between DBAC and TBAC were analyzed by Jackson et al. (2010) and are presented in Table 1.

Table 1. *Dimension-based versus task-based assessment center design features*

Dimension-based	Task-based
Simulation exercises form the backbone of the assessment	Same
Use of multiple assessors	Same
Use of standard-setting training	Same
Pooling of scores by way of clinical discussion or arithmetic integration	Same, except arithmetic pooling is preferred
Based on job analysis	An even greater focus on the importance of job and task analysis
Design focus is on the identification of relevant dimensions	Design focus is on the job-relevance of exercise content
Simulation exercises are used for the assessment of alleged dimensions	Simulation exercises are used for the assessment of work-related behavioral output
Correlated exercise effects interpreted as method bias	Correlated 'exercise effects' interpreted as evidence for general and situationally-specific performance
Scored on the basis of same dimensions aggregated across exercises plus an overall rating	Scored on the basis of behavioral checklists aggregated within exercises plus an overall rating
Intended to tap alleged cross-situationally stable dimensions as output	Intended to tap general performance and situationally-specific elements of performance as output
Formalizes assessment around underlying characteristics	Formalizes assessment around observable behavior
Engages assessors in relatively complex cognitive processes around cross-situational stability	Simply scored within exercises, yet implicitly acknowledges the complexity of behavioral output
Feedback is provided in terms of global dimensions that may succumb to exercise effects	Feedback is provided on the basis of behavioral responses within exercises

Source: Jackson et al., 2010

Debates about the internal consistency of DBAC ratings

A gap between the theoretical MTMM approach applied in DBACs and the available empirical evidence reveals that DBACs have a low internal consistency of ratings across the exercises (Lance et al., 2004; Lievens & Conway, 2001; Sackett & Dreher, 1982). Despite the initial assumption that dimensions were stable cross-contextual constructs, a larger number of recent studies have emphasized that these do not work as they were expected to and that the correlation of the same PEDRs between exercises is actually lower than the correlation of different PEDRs within the same exercise.

The conclusion that dimensions are not (or at least to a certain extent not) equally displayed in all exercises in which they are measured questioned the reliability of DBAC ratings and, implicitly, that of AC methodology (Lance et al., 2000). This idea was overcome by a number of studies that

have shown that the issue regarding reliability of DBAC ratings does not concern a methodological bias (Arthur et al., 2008), but actually the reliable variance in AC ratings is attributed to exercises and to a general performance factor (Jackson et al., 2010, 2016; Putka & Hoffman, 2013).

The current state of knowledge in this domain is thus one of uncertainty: the psychometric properties of TBACs are understudied, but on the other hand DBACs may be saved by models emphasizing broad dimensions, that show better results. It is clear that, based on the issues concerning the internal consistency of AC ratings, it is possible that the TBAC approach may be a better alternative to the traditional DBAC approach, since its measurement strategy is focused on the actual performance in exercises, i.e., the factor that explains the most reliable variance in AC ratings. Nevertheless, other psychometric properties of TBACs are understudied, and their uncritical embracement may be premature (Hoffman et al., 2011). On the other hand, other solutions may improve the internal consistency of DBACs: newer studies have found a better fit for models using broad dimensions, such as broad managerial skill dimension taxonomies (Hoffman et al., 2011; Shore et al., 1990).

The current study

The conclusion that TBAC could be a better alternative to DBAC has been reached from two independent research directions: one that has shown that dimensions used in DBAC are not cross-contextual and the second that has found that reliable variance in ratings is attributed to performance in exercises. These two approaches have not been directly compared so far. The current study addresses this gap and compares the model fit resulting from the two measurement strategies – DBAC and TBAC - on the same data sample.

Research Question: Will a CFA model that relies on TBAC show better fit to the data than a model based on DBAC?

Method

Participants and Procedure

The sample consisted of 126 Romanian students who took part in a developmental AC. In terms of demographics, mean age of participants was 22.28 years (SD = 5.76), 51% of the sample were females and 49% males, 87% undergraduate students and 13% postgraduates. Participants were invited to enroll in an Assessment and Development Center with the help of different organizations such as the College Student Counseling Centers and several NGOs for students, as well as via announcements posted on social media platforms such as Facebook and LinkedIn. During the application process, participants were presented with the purpose of the research and invited to voluntarily enroll in the study. After the enrollment, they received an online invitation for two AC exercises. Both exercises were time limited to three hours. Participants were allowed to solve the exercises at a time of their choosing and responses were collected in writing via an online assessment platform. At the end of the AC, participants received a development report and could opt for a one-on-one feedback session with the assessors.

Assessors

The team of assessors consisted of 16 psychology master students (average age = 24.31 years, *SD* = 2.75, 86% females). Ahead of the actual AC, the assessors participated in two days of frame-of-reference training (Roch et al., 2012), during which they were provided with the relevant details about the methodology and procedure of the AC. Also, the training included information about the dimensions measured, exercises and examples of good and poor performance behaviors, as well as practice sessions.

Variables

For the design of the AC, a taskforce was created, which consisted of three subject matter experts who were senior professionals certified in A&DC methodology. The taskforce analyzed the profiles of top performing students, and consulted the

literature to identify key factors of academic performance. Based on this analysis, we identified that high performing students tended to get involved in extracurricular activities, such as volunteering at conferences or for work in student associations. Therefore, two exercises – analysis and in-tray, were designed to simulate tasks related to extracurricular activities for students (Thornton & Rupp, 2006).

Within the analysis exercise, participants took on the role of the Student Council President at an imaginary university. They were asked to analyze information regarding a master program and propose solutions to improve the program. In the in-tray exercise, participants took on the role of a volunteer coordinator at a conference and were asked to plan and organize the conference.

Two key predictors for academic performance were identified: general mental ability (Kuncel et al., 2004) and conscientiousness (Poropat, 2009). For the DBAC model, two dimensions – problem solving and organizing and planning – were operationalized based on these traits. The dimensions were defined using the taxonomy proposed by Arthur et. al. (2003). The behavioral items are presented in Appendix 1. For TBAC, tasks related to each exercise were operationalized into behavioral indicators. The items are shown in Appendix 2.

After the participants completed the exercises, assessors independently scored dimensions for both the DBAC and TBAC models, directly in the assessment platform. Each behavioral item was rated by two assessors on a 1-5 scale (1 – poor performance to 5 – excellent performance). The final ratings were computed based on an assumption of equal weights (averages).

Materials are openly shared here: https://osf.io/de8vt/?view_only=80fbe722b27840748f987758f18f5ae3

CFA Models

We tested the alternative CFA models using the lavaan package in R (Rosseel, 2012). Consistent with previous research, we fit two different CFA models in order to evaluate the dimensionality underlying the AC ratings as follows: *Model 1* (a *K*-exercise-*J*-dimension model) fit 2 correlated exercise factors and 2 correlated dimension factors, being related to the DBAC approach (Anderson et al., 2006; Bowler & Woehr, 2006; Hoffman et al., 2011), and *Model 2* (a *K*-exercise-one-dimension model) fit 2 correlated exercise factors and one dimension factor, related to TBAC approach (Hoffman et al., 2011; Lance et al., 2004).

In order to test the model fits, we examined the standardized maximum likelihood parameter estimates to determine whether a proper solution had been obtained. We eliminated items that were inconsistent with the data (Byrne, 2011; Hoffman et al., 2011), taking into account the standardized factor loadings and factor correlations. The final models were evaluated based on the chi-square statistic, the root mean squared error of approximation (RMSEA), Bentler's (1990) comparative fit index (CFI) and the Tucker–Lewis index (TLI) (Hu & Bentler, 1999). The two models were compared with each other using Vuong's test (Vuong, 1989).

Results

Table 2 presents the descriptive statistics for the AC dimensions. The average ratings of dimensions ranged from 1.98 (SD = 0.77) to 3.15 (SD = 0.87). The internal consistency was above .90.

Table 3 shows the CFA model fit indexes. After the examination of the standardized maximum likelihood parameter estimates, proper solutions were obtained for both models, with good fit indexes (Hu & Bentler, 1999): *RMSEA* values below .08 (.05; .06) as well as values of $CFI \geq .95$ (.98; .98).

The comparison analysis showed that *Model 2* (TBAC model) fits better to the data than *Model 1* (DBAC model) ($z = -16.15$, $p < .001$), confirming the *Research Question*.

Table 2. *Descriptive Statistics*

Dimension	*M*	*SD*	Alpha Cronbach
Case Study			
Problem Solving	2.64	0.90	.96
Organizing and Planning	1.98	0.77	.91
Task-based Dimension	2.39	0.87	.93
In-basket			
Problem Solving	3.13	0.84	.97
Organizing and Planning	3.15	0.87	.94
Task-based Dimension	2.99	0.87	.91

Table 3. *Confirmatory factor analysis model fit statistics and Vuong's Test*

Model	***Chi-square***	***df***	***p***	***CFI***	***TLI***	***RMSEA***
Dimension-centric model (Model 1)	191.82	146	<.01	.98	.97	.05
Task-based dimension model (Model 2)	48.00	34	<.05	.98	.97	.06
Vuong's Test	z	p				
Model 2 fits better than Model 1	-16.16	<.001				

Discussion

Our study directly compared the two competing assessment strategies of DBAC and TBAC against the same data sample. The results supported the latter approach, indicating better fit indexes for TBAC model.

The TBAC Model

TBAC came out as a promising alternative in reply to the low internal consistency of DBAC ratings. However, despite the fact that the TBAC approach addresses the lack of practical relevance of dimensions in the DBAC model (Jackson et al., 2016), it also brings with it some methodological shortcomings, the main one being in regard to the construct underlying AC ratings.

Construct Validity of TBAC ratings

The construct validity of TBAC ratings is understudied. The general assumption is that they group in a general dimension representing participant performance (Jackson et al., 2016; Jackson et al., 2010); even though TBAC ratings are correlated with other traits, they are not actually assumed to be based on a specific psychological dimension.

Performance in AC exercises is rather seen as a different construct, that varies from one assessment and performance context to another. The whole mechanism is explained through a systematic-deterministic approach where multiple factors concerning participant, organization, exercise design and raters interact and lead to the actual performance that

is displayed in each context (Jackson et al., 2010).

In any case, more evidence is needed to establish whether TBAC should be defined as a psychological construct operationalized by a number of dimensions, or if it is indeed based on one general performance factor. New research could explore whether task-based performance as shown in TBAC ratings may be grouped in taxonomies.

Is performance in TBAC measured appropriately?

Another shortcoming of the TBAC approach is related to the manner in which the performance factor is measured. Since it is viewed as a construct that changes from one context to another, the MTMM principle that a trait (dimension) should be measured by multiple methods (exercises), is not met (Rupp et al., 2015). Further research into the TBAC approach could investigate whether a model in which the same tasks are measured in multiple exercises leads to better results in terms of construct validity, as well as whether same task-performance is cross-contextual.

The DBAC model

Even though the comparison between the two models supported TBAC, good fit indexes were also obtained for DBAC. Other studies indicated low fit indices for DBAC models (Hoffman et al., 2011; Lance et al., 2000). One of the reasons for obtaining a good fit could be related to the parsimony of our model. The model included only two dimensions defined based on the key predictors for academic performance: general mental ability (Kuncel et al., 2004) and conscientiousness (Poropat, 2009). We also point out that the limited number of dimensions may affect the construct validity of the model (Gaugler & Thornton III, 1989).

Dimensions in the DBAC approach are well studied (Arthur et al., 2003; Bowler & Woehr, 2009; Meriac et al., 2008). DBAC models follow the MTMM methodology, attempting to operationalize performance in ACs as comprehensively as possible using multiple dimensions that are measured in multiple exercises. The dimensions are defined based on the ideal profile of the targeted job, focusing on its key psychological traits, which are therefore operationalized into behaviors. It is assumed that psychological traits that are behind each behavior are activated in the exercises. Each situation provides cues that trigger specific traits which lead to the manifestation of behaviors and, thus, to performance in exercises (Lievens et al., 2006).

The drawback of the DBAC approach is related to the traits underlying the dimensions, which are differently activated from one situation to another as well as to the fact that different traits may interact within the same context and overlap their effects in the manifestation of the behaviors (Haaland & Christiansen, 2002). Nevertheless, some studies indicated that models with dimensions spanning broad factors turned out to be a better fit to the data (Hoffman et al., 2011). DBAC designs based on broad performance factor models such as Shore's (1990) may lead to superior results in terms of both construct validity and the internal reliability of AC dimension ratings.

Limitations

One limitation of the study is related to the sample that consists of students, most of them undergraduates. The participants took part in ACs for research purposes and were provided with the development opportunity to receive a one-on-one feedback session. The models were not tested on data collected from real ACs used in selection or development processes for employees and, thus, this might limit the extent to which the conclusions of this study may be generalized to employee samples.

A second limitation concerns the number of dimensions used for testing the models, which may influence the construct validity of the models (Gaugler & Thornton, 1989). For instance, multiple dimensions included in the TBAC model could have reduced its parsimony and weakened the fit indexes, while for a DBAC model more dimensions could have led to obtaining unacceptable fit indexes (Hoffman et al., 2011; Lance et al., 2000).

A third limitation is posed by the fact that the models were tested only in written

exercises. The assessors only evaluated responses submitted by the participants, without having the opportunity to prompt behaviors, as it usually happens during interactive, "live" exercises. This hampers the generalizability of our conclusions to other types of AC exercises.

Practical Implications and Future Research

The results of our study revealed superior fit indices for a TBAC model when directly compared to DBAC, supporting a task-oriented measurement strategy in ACs (Hoffman et al., 2015; Jackson et al., 2010, 2016). Despite promising results, more evidence is needed to move from a traditional dimension-based assessment to a task performance-based approach.

For future research, we recommend replicating this study on data collected from ACs conducted in selection/promotion or development processes in organizations to investigate whether the same results are achieved in real-life situations. The research design could be expanded by defining models that include a wider range of competencies, tasks and types of exercises derived from job analyses of roles within organizations.

Another issue worth considering for future research is the criterion-related validity of TBAC models compared to DBAC. By showing higher internal consistency, it is assumed that TBAC ratings might predict future job performance better than DBAC ratings. However, the criterion-related validity of TBAC models has not been widely studied and little is known about how behavioural ratings are related to future outcomes. New studies in this area may highlight differences, such as the fact that TBACs better predict task performance, while DBACs better predict contextual performance (Motowildo, Borman & Schmit, 2009).

Finally, more evidence is needed to fully understand what constitutes performance in the TBAC approach and whether certain psychological traits underlie it. To integrate MTMM principles into TBAC design, we recommend testing models that include homogeneous task dimensions (e.g., customer relationships) and evaluating them in multiple contexts (e.g., two role-plays both simulating customer interactions). Alternatively, DBAC internal consistency issues could be addressed by using broad dimensions models derived from established performance taxonomies such as Shore's (1990) and, thus, a mixed approach between the two models, DBAC and TBAC, could lead to superior results in terms of internal consistency, construct validity and criterion-related validity.

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

Arthur, W., Day, E. A., Mcnelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology*, *56*(1), 125–153. https://doi.org/10.1111/j.1744-6570.2003.tb00146.x

Arthur, W., Day, E. A., & Woehr, D. J. (2008). Mend it, don't end it: An alternate view of assessment center construct-related validity evidence. *Industrial and Organizational Psychology*, *1*(1), 105–111. https://doi.org/10.1111/j.1754-9434.2007.00019.x

Bentler, P. M. (1990). Comparative fit indexes in structural models. Psychological Bulletin, *107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

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

Bowler, M. C., & Woehr, D. J. (2009). Assessment center construct-related validity: Stepping beyond the MTMM matrix. *Journal of Vocational Behavior*, *75*(2), 173–182. https://doi.org/10.1016/j.jvb.2009.03.008

Bycio, P., Alvares, K. M., & Hahn, J. (1987). Situational specificity in assessment center ratings: A confirmatory factor analysis. Journal of Applied Psychology, *72*(3), 463–474. https://doi.org/10.1037/0021-9010.72.3.463

Cahoon, M. V., Bowler, M. C., & Bowler, J. L. (2012). A reevaluation of assessment center construct-related validity. *International Journal of Business and Management*, *7*(9), p3. https://doi.org/10.5539/ijbm.v7n9p3

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81-05. https://doi.org/10.1037/h0046016

Fleenor, J. W. (1996). Constructs and developmental assessment centers: Further troubling empirical findings. *Journal of Business and Psychology*, *10*(3), 319–335. https://doi.org/10.1007/BF02249606

Gaugler, B. B., & Thornton, G. C. (1989). Number of assessment center dimensions as a determinant of

assessor accuracy. *Journal of Applied Psychology, 74*(4), 611–618. https://doi.org/10.1037/0021-9010.74.4.611

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

Haaland, S., & Christiansen, N. D. (2002). Implications of trait-activation theory for evaluating the construct validity of assessment center r atings. *Personnel Psychology, 55*(1), 137–163. https://doi.org/10.1111/j.1744-6570.2002.tb00106.x

Hoffman, B. J., Melchers, K. G., Blair, C. A., Kleinmann, M., & Ladd, R. T. (2011). Exercises and dimensions are the currency of assessment centers. *Personnel Psychology, 64*(2), 351–395. https://doi.org/10.1111/j.1744-6570.2011.01213.x

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

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

Jackson, D. J. R., & Englert, P. (2011). Task-Based Assessment Centre Scores and their Relationships with Work Outcomes. *New Zealand Journal of Psychology, 40*(2), 11.

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

Kleinmann, M., & Ingold, P. V. (2019). Toward a Better Understanding of Assessment Centers: A Conceptual Review. *Annual Review of Organizational Psychology and Organizational Behavior, 6*(1), 349–372. https://doi.org/10.1146/annurev-orgpsych-012218-014955

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic Performance, Career Potential, Creativity, and Job Performance: Can One Construct Predict Them All? *Journal of Personality and Social Psychology, 86*(1), 148–161. https://doi.org/10.1037/0022-3514.86.1.148

Kuncel, N. R., & Sackett, P. R. (2014). Resolving the assessment center construct validity problem (as we know it). *Journal of Applied Psychology, 99*(1), 38–47. https://doi.org/10.1037/a0034147

Lance, C. E. (2008). Why Assessment Centers Do Not Work the Way They Are Supposed To. *Industrial and Organizational Psychology, 1*(1), 84–97. https://doi.org/10.1111/j.1754-9434.2007.00017.x

Lance, C. E., Lambert, T. A., Gewin, A. G., Lievens, F., & Conway, J. M. (2004). Revised Estimates of Dimension and Exercise Variance Components in Assessment Center Postexercise Dimension Ratings. *Journal of Applied Psychology, 89*(2), 377–385. https://doi.org/10.1037/0021-9010.89.2.377

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

Lievens, F., De Koster, L., & Schollaert, E. (2008). Current theory and practice of assessment centers: the importance of trait activation. In S. Cartwright & C. L. Cooper (Eds.), *Oxford Handbook of Personnel Psychology* (pp. 215–233). https://doi.org/10.1093/oxfordhb/9780199234738.003.0010

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., & Conway, J. M. (2001). Dimension and exercise variance in assessment center scores: A large-scale evaluation of multitrait-multimethod studies. *Journal of Applied Psychology, 86*(6), 1202–1222. https://doi.org/10.1037/0021-9010.86.6.1202

Lievens, F., & Thornton, G. C. (2017). Assessment Centers: Recent Developments in Practice and Research. In A. Evers, N. Anderson, & O. Voskuijl (Eds.), *The Blackwell Handbook of Personnel Selection* (pp. 243–264). Blackwell Publishing Ltd. https://doi.org/10.1002/9781405164221.ch11

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

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. Psychological Bulletin, *135*(2), 322–338. https://doi.org/10.1037/a0014996

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., Woehr, D. J., Mishra, V., & Kieszczynska, U. (2012). Rater training revisited: An updated meta-analytic review of frame-of-reference training: Rater training revisited. *Journal of Occupational and Organizational Psychology, 85*(2), 370–395. https://doi.org/10.1111/j.2044-8325.2011.02045.x

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rupp, D., Hoffman, B., Bischof, D., Byham, W., Collins, L., Gibbons, A., Hirose, S., Kleinmann, M., Kudisch, J., Lanik, M., Jackson, D., Kim, M., Lievens, F., Meiring, D., Melchers, K., Pendit, V., Putka, D., Povah, N., Reynolds, D. & Thornton, G. (2015). Guidelines and ethical considerations for assessment center operations. *Journal of Management, 41*(4), 1244–1273. https://doi.org/10.1177/0149206314567780

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

Shore, T. H., Thornton, G. C., & Shore, L. M. (1990). Construct validity of two categories of assessment

center dimension ratings. *Personnel Psychology*, *43*(1), 101–114. https://doi.org/10.1111/j.1744-6570.1990.tb02008.x

Schmidt, F. L., & Hunter, J. E. (2016). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. In G. J. Boyle, J. G. O'Gorman, & G. J. Fogarty (Eds.), *Work and organisational psychology: Research methodology; Assessment and selection; Organisational change and development; Human resource and performance management; Emerging trends: Innovation/globalisation/technology* (pp. 1–27). Sage Publications, Inc.

Thornton, G. C. III, & Rupp, D. E. (2006). Assessment centers in human resource management: Strategies for prediction, diagnosis, and development. Lawrence Erlbaum Associates Publishers.

Vuong, Q.H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, *57* (2), 307-333

Appendix 1

DBAC Behavioral Items Description

Problem Solving

Negative Behaviors	Positive Behaviors
Misses information when analyzing, takes into account a limited number of available data.	Analyzes a large amount of information available.
Does not generate ideas or solutions for existing problems.	Generates multiple ideas or solutions to solve existing problems.
Does not look for new information, relies on existing information when solving problems.	Search for new information in order to better understand the existing problem.
Does not have a clear understanding, misinterprets information from the brief.	Has a clear understanding of the information provided, understands the context of the exercise in its complexity.
Mixes up essential information with irrelevant details.	Identifies the essential elements
Does not propose actions to solve problems, adopting a passive approach.	Selects a course of action in order to effectively solve problems.
Does not efficiently use existing resources to solve problems.	Looks for new ways to use existing resources to solve problems in the most efficient way.
Does not generate imaginative solutions.	Generates and recognizes imaginative solutions.

Organizing and Planning

Negative Behaviors	Positive Behaviors
Approaches tasks in a disorderly, unstructured manner.	Has a systematic and orderly approach to tasks.
Does not efficiently organize the work of others in order to ensure the fulfillment of tasks.	Effectively organizes the work of others in order to fulfill the tasks.
Does not anticipate factors that could appear along the way and impact the achievement of objectives.	Anticipates factors that could intervene along the way.
Does not keep track of activities.	Keeps track of activities.
Does not effectively prioritize activities.	Effectively prioritizes activities.
Does not establish a timeline of activities.	Establishes a timeline of activities.
Focuses on immediate problems, makes short term plans.	Plans long-term activities.

Appendix 2

TBAC Behavioral Items Description

Analysis Exercise

Negative Behaviors	Positive Behaviors
Is not aware of the strengths and weaknesses of the master's program, does not take into account the results from the university rankings.	Highlights the strengths and weaknesses of the master's program.
Does not analyze the information from the student satisfaction survey nor the typology of the master's students.	Effectively uses the information from the student satisfaction survey, and other information/trends.
Does not analyze the revenues which could be obtained by transforming the current master's program.	Calculates how much the income would increase and what would be the financial benefits that could be obtained by implementing the new initiative.
Is unspecific about the master's initiative, does not have a clear understanding of the context.	Effectively identifies the advantages and disadvantages of the options for the master's program; makes well-reasoned recommendations regarding new initiatives.
Does not address the topic of the online platform.	Proposes solutions to effectively manage logistics costs; makes recommendations regarding the e-learning platform.
Does not take into account the needs and profile of the teachers when making recommendations.	Proposes actions to address teachers' needs; takes into account their typology when making recommendations.

In-tray Exercise

Negative Behaviors	Positive Behaviors
Does not address the issue about keynote speakers (considers little information about keynote speakers, offers vague recommendations).	It presents clear recommendations regarding the choice of keynote speakers, offering arguments based on a comprehensive analysis.
Does not propose recommendations for choosing the topic; the implications of the topics chosen are not clear to them.	Demonstrates that is aware of the implications of the topics chosen.
Does not realize that the budget is exceeded; proposes vague solutions.	Calculates exactly how much the budget is exceeded, proposes solutions to accommodate the needs of the project.
Does not address the topic of the promoting campaign.	Comes up with new ideas and a plan for the promoting campaign.
Is not concern of the project team.	Proposes actions to organize the project team; considers the needs of the project, looking for suitable people for the new roles.