

Running head: DYSFUNCTIONAL ELEMENTS OF ASSESSMENT CENTERS

The Perceived Nature and Incidence of
Dysfunctional Assessment Center Features and Processes

Dr. C. Dewberry and Dr. J.R. Jackson

Birkbeck, University of London

Published in the International Journal of Selection and Assessment, 2016

Abstract

The aims of this research are to identify, for the first time, the dysfunctional features and processes perceived to take place in assessment centers (ACs) from multiple perspectives (assessment center designers, assessors, and candidates) and to indicate the frequency of these phenomena. Two surveys were conducted in this study. In the first, a wide variety of dysfunctional processes and events were identified, and, in the second, many of these processes are reported to occur with regularity. Based on these findings, it is proposed that ACs should be construed, researched, and managed, not only as large-scale psychometric systems, but also as complex administrative, social, and political events susceptible to a broad range of dysfunctional phenomena.

DRAFT

The Perceived Nature and Incidence of Dysfunctional Assessment Center Characteristics

During the last 20 years several large-scale surveys of assessment centers (ACs) have been undertaken in North America, Western Europe, and South Africa. As a consequence, much is known about the nature and incidence of AC structures and practices, including the steps taken in developing ACs, the nature of dimensions and exercises, assessor training, and data integration techniques (Eurich, Krause, Cigularov, & Thornton, 2009; Krause & Gebert, 2003; Krause, Rossberger, Dowdeswell, Venter, & Joubert, 2011; Krause & Thornton III, 2009; Lowry, 1996; Spychalski, Quiñones, Gaugler, & Pohley, 1997).

As well as clarifying the strengths of real-world AC design and implementation, these surveys also reveal a number of problems. These problems are in addition to the much-studied construct validity issue (Fleenor, 1996; Lance et al., 2000; Russell & Domm, 1995; Shore, Shore, & Thornton, 1992), with which this article is not concerned. For example, after surveying AC practices in the United States, Eurich et al. (2009) concluded that the majority of organizations running them did not train assessors for long enough. Awareness of the potential weaknesses in the design of some ACs has led to suggestions as to how they might be improved, with Caldwell, Thornton and Gruys (2003) drawing attention to 10 “classic” problems to avoid in the design of ACs (e.g. weakly defined dimensions, poor exercises, and the use of unqualified assessors), and Krause and Thornton (2009) suggesting that AC designers should conduct pilot tests of exercises before they are implemented, integrate additional diagnostic procedures, consider the diversity of assessors, increase observer training, and evaluate AC performance.

These observations and recommendations are potentially helpful to AC designers. However, they do not provide detailed information about the nature of the problems currently occurring in ACs, the frequency of these problems, or the steps which might be taken to address them. For example, whilst inadequate assessor training is cited as a problem by Caldwell et al. (2003), and an increase in the length of assessor training is recommended by Krause and Thornton (2009), neither sets of authors document the ways in which current training is inadequate or provide guidance about the content of the additional training they recommend. Such guidance is restricted, at least in part, because the results of AC surveys (Eurich, et al., 2009; Krause & Gebert, 2003; Krause, et al., 2011; Krause & Thornton III, 2009; Lowry, 1996; Spsychalski, et al., 1997) generate limited information about the dysfunctional social, psychological, and administrative features and processes which may occur in them.

There are at least four reasons for the absence of detailed information about dysfunctional AC characteristics and processes. First, almost all published AC surveys were designed to provide “stock-checks” of the general structural features and processes used in ACs and do not focus specifically on the features and processes which are problematic and dysfunctional. Second, the questionnaires used in these surveys generally focus on the “hard” structural features and processes of the centers (e.g. the number of dimensions against which candidates are assessed, and the length of assessor training) rather than “soft” psychological and social processes which may undermine AC efficacy (e.g. the extent to which junior assessors may defer to senior ones in consensus meetings [Dewberry, 2011]). Third, the individuals completing the surveys (typically HR managers) are de facto providing an official record of the features of ACs in the organization employing them, and, as a consequence, may feel constrained in the way they report inadequacies and deficiencies in these ACs. Fourth, although these individuals (typically HR managers) possess factual

information about the AC upon which they are reporting, they may lack the detailed knowledge of those who have direct experience of them as AC designers, assessors, and candidates.

An exception to the general AC surveys discussed above is one specifically designed to examine AC problems (Lievens & Goemaere, 1999). Lievens and Goemaere interviewed a member of the personnel department in 23 Belgian organizations. These human resource professionals viewed five areas as problematic: (a) exercise selection and development, (b) assessor selection and training, (c) observation and rating processes, (d) issues related to the overall assessment rating, and (e) feedback to candidates and management. Although, unlike other surveys of ACs, their research focused directly on AC problems, Lievens and Goemaere's data represented one perspective only (personnel managers), and the findings, presented in the context of a professional forum article, were described informally (e.g. "development costs and the lack of expertise to develop new exercises topped the comments received" [p. 216]).

The purpose of the present article is to extend theory and research on the dysfunctional social, administrative, and political features of ACs in four ways. First, information about the perceived dysfunctional characteristics of ACs is sought from three distinct perspectives: designers, assessors, and candidates. Second, unlike previous studies, respondents provide information anonymously and not as a representative of a particular organization. Third, respondents are requested to list any AC problems that they have observed in their own words, without any interaction with the researcher or any pre-conceived list or classification of what these problems might be. Fourth, the frequency with which designers, assessors, and candidates perceive these problems to occur is examined.

AC designers and assessors are required to hold expertise relevant to their positions. The views of subject matter experts (SMEs) are widely used as a source of research evidence

in applied psychology (Dries, Pepermans, & Carlier, 2008; Flores, Zheng, Rau, & Thomas, 2012; Francis-Smythe, Haase, Thomas, & Steele, 2013; Jones & Born, 2008; Motowidlo & Beier, 2010; Sebastian, Ramos, Stumbo, McGrath, & Fairbrother, 2014; Woehrmann, Deller, & Wang, 2014). Although AC assesses are not organizational SMEs in the usual sense of the term, they have relevant experience of ACs from a distinctive perspective, and their perceptions may impact in several ways on organizations using ACs, including the likelihood that they will accept a job offer and their first impression of the organization.

Two studies are reported. In Study 1, AC designers, assessors, and candidates describe any AC characteristics and processes they have observed which may have undermined the quality of the center (designers and assessors) or which were inappropriate, hard to justify, upsetting to others involved in the AC process, or wrong (candidates). In Study 2 the frequency with which these characteristics and processes are observed is examined. Both studies were specifically concerned with assessment centers (used for staff selection), and not development centers (used for staff development).

Study 1

Method

Participants. Members of a British Psychological Society Working Party on AC standards assembled by the British Psychological Society in 2013 contacted, by email, people they were aware of with experience of ACs in the UK as designers, assessors, or candidates. Those who responded, after answering the survey, were invited to circulate a website http address among others with experience of ACs. The survey was completed by 32 AC designers (mean age = 43 years, 60% female), 73 candidates (mean age = 40 years, 32% female), and 68 assessors (mean age = 44 years, 49% female).

Instrument. An online questionnaire, with separate sections for those who had been AC designers, assessors, and/or candidates, was presented to participants. The instructions at the beginning of each section are shown in Appendix 1. Beneath each set of instructions text boxes were available for participants to describe each of the problematic phenomena they had observed.

Results and Discussion

Duplicate problems and those which could not be directly observed (e.g. “minority influence” and “groupthink”) were eliminated and a final list of AC problems was established for the AC designer, assessor, and candidate roles. Where necessary, the description of a problem was re-worded to improve clarity and concision. The final lists of problems derived from designers is shown in Table 1, from assessors in Table 2 (for pre-consensus meeting problems) and Table 3 (for consensus meeting problems), and from candidates in Table 4. AC designers identified the smallest number of problems (30), with candidates (43), and especially assessors (93), specifying substantially more.

Two raters examined the tables, and after discussion agreed that the problems could be usefully and meaningfully categorized as those associated with (a) AC design; (b) the implementation of ACs after they have been designed; (c) assessor biases, errors, and poor behavior; (d) training deficiencies; (e) lack of consideration for candidates; and (f) others. These raters then independently placed each of the problems in one of these six categories. The Gower statistic (Gower, 1971) was used to examine the degree of agreement between the raters. Relative to the maximum possible absolute discrepancy between them, the observations agreed to 95% of each other's values. Any remaining points of disagreement were resolved by discussion and a final set of categorized problems was thereby obtained. The category into which each item was placed is shown in Tables 1 to 4.

The frequency and percentage of each category of problem for designers, assessors, and candidates is shown in Table 5. A chi square analysis (excluding the ‘training’ and ‘other’ categories to ensure the percentage of cells with expected frequencies less than 5 was minimized) indicated that the proportion of problems specified in the AC design, AC implementation, assessor-related, and candidate consideration categories differs across designers, assessors and candidates, $\chi^2 (6) = 105.1, p < .001$.

Insert Tables 1 to 5 about here

The large number of AC problems identified here add substantially to those specified previously (Caldwell, et al., 2003; Krause & Thornton III, 2009; Lievens & Goemaere, 1999). Former articles generally focus on issues perceived from the perspective of AC designers, such as inadequate job analysis, weakly-defined dimensions, poor exercises, the absence of pre-test evaluations, and unqualified assessors (Caldwell, et al., 2003; Lievens and Goemaere, 1999), and suggest improvements based on these designer-orientated issues (e.g. conduct pilot tests of exercises before they are implemented, integrate additional diagnostic procedures, consider the diversity of assessors, increase observer training, and evaluate the performance of the ACs).

The absence of many of the perceived problems identified by assessors and candidates here from the list of problems identified by Caldwell Thornton and Gruys (2003), Krause and Thornton (2009), Lievens and Goemaere (1999), and by the AC designers in this study, suggests that designers may not be aware of many of the dysfunctional and problematic AC features and processes observed by assessors and candidates.

Whereas over 80% of the problems identified by AC designers were concerned with AC design, and, to a lesser extent, AC implementation, about two thirds of the problems identified by the assessors were concerned with errors, biases and poor behavior they had observed in other assessors. For candidates, the most frequent type of problem was a lack of candidate consideration. These findings strongly suggest that, in order to obtain a comprehensive understanding of the dysfunctional characteristics of ACs, it is necessary to examine them from several perspectives.

Study 2

Study 1 yielded an extensive list of the dysfunctional characteristics of ACs from the perspective of AC designers, assessors, and candidates. The purpose of the second study was to extend this by gathering information about the frequency with which those with experience of ACs perceive these dysfunctional characteristics to occur.

Participants and Instruments

As for Study 1, members of a British Psychological Society working party on AC standards emailed people they were aware of with experience of ACs in the UK with the link to the online survey. In addition, an http link was placed on the British Psychological Societies' Division of Occupational Psychology website, and messages were also placed in two AC and selection focused discussion groups on the *Linked In* website, inviting people with experience of being AC designers, assessors or candidates in the UK to take part. These contacts were asked to pass the http link on to others.

Separate online questionnaires were used for AC designers, assessors, and candidates. Designers were presented with the list of AC features and processes set out in Table 1, assessors with those set out in Tables 2 and 3, and candidates with those set out in Table 4. In

all cases, AC issues were presented in random order. Respondents were asked to estimate the frequency with which they had observed each issue using the scales described below.

Designers. Twenty-five people with experience of designing one or more ACs in the UK completed the questionnaire online. The mean age of the designers was 40 years and 18 (72%) were female. The number of ACs designed by the participants had a strong negative skew (median = 15, interquartile range 5.5 - 47.5). The participants were presented with the 30 AC processes and features derived shown in Table 1. They were asked to indicate how many ACs they had observed each issue in.

Assessors. The questionnaire for AC assessors was completed by 95 people with a mean age of 46 years, 69 (62%) were female. The number of ACs attended in the UK by these assessors had a very strong positive skewed with a median of 100 and an interquartile range 24 to 256. Participants were presented with the 93 AC features and processes shown in Tables 2 and 3. It was anticipated that most assessors would have performed this role a large number of times, and, as a consequence, it would not be possible for them to reliably recall the number in which they had observed each problem or process. Consequently they were asked to indicate how often they had experienced each issue on the following scale: never, rarely, occasionally, quite often, very often, or always or almost always.

Candidates. The candidate section of the questionnaire was completed by 64 participants with a mean age of 40, 35 (55%) were female. The median number of ACs attended in the UK by these candidates was 4 (minimum 2, maximum 12). Participants were presented with the 30 issues shown in Table 3 in random order. As for the designers, they were asked to indicate in how many ACs each problem had been experienced.

Results and Discussion

The percentage of times each designer had experienced each problem was obtained by dividing the number of ACs in which he/she reported experiencing a problem by the total

number of ACs that he/she has designed. The mean of these percentages across all designers is shown in Table 1. The proportion of assessors indicating that each problem occurred either “quite often”, “very often” or “always or almost always” in Tables 2 and 3. Table 2 shows problems which occur outside of consensus meetings (or “wash-ups”). Table 3 is concerned only with issues which arise in these meetings. For candidates, the percentage of times each candidate reported experiencing each problem was obtained by dividing the number of ACs in which he/she reported experiencing the problem by the total number of ACs in which he/she had been assessed. The mean of these percentages across all candidates is shown in Table 4. In all four tables the AC features and processes to which participants responded are presented in order of their reported frequency (high to low).

To examine whether the relative frequency of AC problems observed by assessors outside the consensus meeting (“never” to “always or almost always”) is predicted by the number of ACs they estimated experiencing, the number of these which were “in house”, and the number in the private sector, a standard multiple regression analysis was carried out. The regression model was not significant $F(3, 97) = 0.72, p > .05$. A second standard multiple regression using the same three independent variables to predict the frequency of AC problems observed by assessors in consensus meetings was also non-significant $F(3, 72) = 0.85, p > .05$. This suggests that the perceived (relative) frequency of AC problems by assessors does not differ as a function of assessor experience, public versus private sector organizations, or in-house versus externally designed ACs.

The proportion of private sector (versus public sector) ACs attended by candidates was significantly correlated with the total number of problems they observed. $r=.26$, $p=.04$. Candidates were therefore more likely to observe problems when ACs were run in the private sector. No inferential statistical analyses were undertaken on the problems reported by AC designers due to the restrictive number of designers sampled.

General Discussion

Taken together, the two studies reported here provide, for the first time, an extensive list of the dysfunctional characteristics observed to occur in ACs from the perspective of designers, assessors, and candidates, and an index of the frequency with which these characteristics are observed to occur. The findings compliment and extend the limited literature on AC problems (Caldwell, et al., 2003; Lievens & Goemaere, 1999) by identifying, empirically, the multifaceted nature and frequency of perceived AC problems from three distinct perspectives: AC designers, assessors, and candidates, and by enabling respondents to answer questions anonymously rather than to respond as a representative of a particular organization running ACs. A substantial number of dysfunction characteristics were identified by designers (30), candidates (42), and especially assessors (92); and as Table 5 shows the nature of these characteristics varied across these groups, with AC designers predominantly drawing attention to design issues, assessors to biases and errors in the assessment process and poor assessor behavior, and candidates to ways in proper consideration is not shown to them.

Theoretical Contribution

To date the principal theoretical debate in relation to ACs has concerned their construct validity, with about 50 articles on this issue (Guenole, Chernyshenko, Stark, Cockerill, & Drasgow, 2013), including several meta-analyses (Bowler & Woehr, 2006;

Lance, Lambert, Gewin, Lievens, & Conway, 2004; Lievens & Conway, 2001; Woehr & Arthur, 2003). Within the context of this debate, ACs are implicitly construed as large-scale psychometric measures and variance in AC ratings is decomposed into identifiable sources, particularly exercises and dimensions, and sometimes also assessors, candidates, and interaction effects (for a recent example, see Putka and Hoffman [2013]). In furthering an understanding of the reliability and validity of ACs, these studies are clearly of considerable theoretical importance. In contrast, the current study aims to address the relatively under-researched topic of theoretical value: the dysfunctional features of ACs which may undermine their performance and their perceived utility as perceived from the distinctive perspectives of AC designers, assessors, and candidates.

The results of the surveys conducted here reveal a rich and comprehensive set of perceived problems with ACs. They indicate that the broad range of problems to which designers, assessors and candidates draw attention are concerned, not only with psychometric assessment directly, but also with a complex set of social, administrative, financial, and political issues. The processes and structures underpinning these issues are interlocking and mutually-influencing. For example, an AC designer may rationally decide to use multiple assessors for each candidate in an AC on the grounds that this will help to minimize the influence of systematic differences in the tendency of assessors to be severe versus lenient in their assessment of candidates (Bernardin, Cooke, & Villanova, 2000; Bernardin, et al., 2009; Kane, et al., 1995). However, the introduction of multiple assessors will affect the AC in other areas. It will increase administrative complexity (e.g. there are more assessors to train, and a more complex assessment timetable will be required), social complexity (assessors will have more opportunities to discuss candidates informally as well as formally, may exchange more information about candidates with other assessors, with a concomitant increase in social biases such as stereotyping and majority and minority influence effects), cost (which may

mean a reduction of resources for other elements of the center), and possibly political complexities (a client organization may wish to ensure that their influence on AC outcomes is not diminished). From a theoretical perspective, it is therefore important that researchers and practitioners construe ACs not only as a measurement and assessment process, but also as a complex set of social, administrative, political, and financial systems in which modifications in one area are likely to have multiple consequences on other areas, many of them difficult to predict, and from the perspective of designers, difficult to detect.

A further contribution of this article is to provide a number of potential explanations for the substantial evidence that the future job performance of AC candidates is better predicted by the weighted or unweighted addition of candidate dimension or exercise scores gathered in ACs than by overall assessment ratings often derived from consensus meeting discussions (Dewberry, 2011; Dilchert & Ones, 2009). Assessor perceptions (see Table 3) that sometimes insufficient time is given to consensus meetings; that discussion can be unfocused or disproportionately influenced by assessors who are senior, “in-house” or have a dominant personality; that assessors actively frame evidence about candidates or place more emphasis on defending their own evaluations than listening to others, as well as a variety of other events of this type, provide a rich source of information with which to develop and test explanations for the relatively poor criterion-related predictiveness of overall AC ratings.

Practitioner Implications

The substantial number of perceived problems with ACs listed in Tables 1 to 4 provides practitioners with a novel checklist of issues to reflect on and to address when designing, trialing, and running ACs. Because it is difficult or impossible for AC designers to directly observe many of these problems (e.g. those set out in Tables 3 and 4), AC designers may find it helpful to systematically gather feedback from assessors and candidates on these issues

when piloting and running ACs, and, in the light of this information, to adjust the design of ACs to counter them where possible.

Practitioners and researchers need to consider the implications of changes and innovations bearing in mind not just the implications at the level of analysis it is introduced but at other levels too. For example, multiple assessors might be a good idea at the assessor level of analysis for exercise assessment reliability and validity. But the explicit and latent implications for this at other levels of analysis need to be considered also. For example, at the interpersonal level of analysis, an increase in the number of assessors may be associated with an increase in the incidence of informal discussion about candidates and, at the intergroup level of analysis, there may be an increase in majority influence effects in consensus meetings.

The observed problems with ACs identified here also provide a useful resource for organizations developing or amending guidelines for ACs, such as the International Task Force on Assessment Center Standards (2015), and the British Psychological Society Division of Occupational Psychology (2015). Furthermore, the problems associated with consensus meetings (see Table 4), when combined with the evidence suggesting that overall AC ratings (often derived from consensus meeting discussions), are less predictive than arithmetic methods of data integration (Dewberry, 2011; Dilchert & Ones, 2009), may help to stimulate a constructive debate about whether consensus meetings should continue to be recommended as a suitable process for data integration, and, if so, what steps can be taken to ensure that any benefits they may bring to the data integration process are not outweighed by the disadvantages arising from the potential problems set out in Table 3.

Limitations

There are two main limitations of this research. First, it is important to recognize that this article is concerned with the *perceived* nature and incidence of problems in ACs rather

than with the *actual* existence of these problems or their *actual* frequency of occurrence. Second, whether some of the problems which participants were asked to report on in Study 2 can be reliably observed (e.g. “Assessors scoring candidates by comparing them to each other, rather than according to pre-specified criteria”) is questionable. Estimates of the perceived incidence of such marginally observable phenomena should not be used as reliable indicator of their actual incidence.

Future research

The results of this study can help to inform research into the steps which might be taken to improve ACs. In particular it would be helpful to know the extent to which these processes actually take place in ACs, whether they tend to be concentrated in some ACs with others relatively free of them, and the how much damage they do to the reliability and criterion-related validity of ACs. It would also be useful to clarify what steps can be taken to eliminate as many of these dysfunctional processes as possible.

References

- Bernardin, H. J., Cooke, D. K., & Villanova, P. (2000). Conscientiousness and agreeableness as predictors of rating leniency. *Journal of Applied Psychology, 85*, 232-236.
- Bernardin, H. J., Tyler, C. L., & Villanova, P. (2009). Rating level and accuracy as a function of rater personality. *International Journal of Selection and Assessment, 17*, 300-310.
- 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. doi: 10.1037/0021-9010.91.5.1114|issn 0021-9010.
- British Psychological Society Division of Occupational Psychology. (2015). The design and delivery of assessment centres. (2015). Leicester, British Psychological Society.
- Caldwell, C., Thornton, G. C., & Gruys, M. L. (2003). Ten classic Assessment Center errors: Challenges to selection validity. *Public Personnel Management, 32*, 73-88.
- Dewberry, C. (2011). Integrating Candidate Data: Consensus or Arithmetic? . In N. Povah & C. Thornton (Eds.), *Assessment and Development Centres: Strategies for Global Talent Management*. Farnham, Surrey: Gower.
- Dilchert, S., & Ones, D. S. (2009). Assessment Center Dimensions: Individual differences correlates and meta-analytic incremental validity. *International Journal of Selection and Assessment, 17*, 254-270.
- Dries, N., Pepermans, R., & Carlier, O. (2008). Career success: Constructing a multidimensional model. *Journal Of Vocational Behavior, 73*, 254-267. doi: 10.1016/j.jvb.2008.05.005
- Eurich, T. L., Krause, D. E., Cigularov, K., & Thornton, G. C. (2009). Assessment centers: current practices in the United States. *Journal of Business and Psychology, 24*, 387-407. doi: 10.1007/s10869-009-9123-3

- Fleenor, J. W. (1996). Constructs and developmental assessment centers: Further troubling empirical findings. *Journal of Business and Psychology, 10*, 319-335.
- Flores, L. G., Zheng, W., Rau, D., & Thomas, C. H. (2012). Organizational Learning: Subprocess Identification, Construct Validation, and an Empirical Test of Cultural Antecedents. *Journal Of Management, 38*, 640-667. doi: 10.1177/0149206310384631
- Francis-Smythe, J., Haase, S., Thomas, E., & Steele, C. (2013). Development and Validation of the Career Competencies Indicator (CCI). *Journal of Career Assessment, 21*(2), 227-248. doi: 10.1177/1069072712466724
- Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics, 27*, 857-871. doi: 10.2307/2528823
- Guenole, N., Chernyshenko, O. S., Stark, S., Cockerill, T., & Drasgow, F. (2013). More than a mirage: A large-scale assessment centre with more dimension variance than exercise variance. *Journal of Occupational and Organizational Psychology, 86*, 5-21. doi: 10.1111/j.2044-8325.2012.02063.x
- International Taskforce on Assessment Center Guidelines. (2015). Guidelines and ethical considerations for assessment center operations *Journal of Management, 41*, 1244–1273. doi: 10.1177/0149206314567780
- Jones, R. G., & Born, M. P. (2008). Assessor constructs in use as the missing component in validation of assessment center dimensions: A critique and directions for research. *International Journal of Selection And Assessment, 16*, 229-238. doi: 10.1111/j.1468-2389.2008.00429.x
- Kane, J. S., Bernardin, H. J., Villanova, P., & Peyrefitte, J. (1995). Stability of rater leniency - 3 studies. *Academy of Management Journal, 38*, 1036-1051.

- Krause, D. E., & Gebert, D. (2003). A comparison of assessment center practices in organizations in German-speaking regions and the United States. *International Journal of Selection and Assessment, 11*, 297-312.
- 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*, 262-275. doi: 10.1111/j.1468-2389.2011.00555.x
- Krause, D. E., & Thornton, G. C. (2009). A Cross-Cultural Look at Assessment Center Practices: Survey Results from Western Europe and North America. *Applied Psychology-An International Review-Psychologie Appliquee-Revue Internationale, 58*, 557-585. doi: 10.1111/j.1464-0597.2008.00371.x
- Krause, D. E., & Thornton III, G. C. (2009). A Cross-Cultural Look at Assessment Center Practices: Survey Results from Western Europe and North America. *Applied Psychology: An International Review, 58*, 557-585.
- Lance, C. E., Lambert, T. A., Gewin, A. G., Lievens, F., & Conway, J. M. (2004). Revised estimates of dimension and exercise variance components in assessment center postexercise dimension ratings. *Journal of Applied Psychology, 89*, 377-385.
- Lance, C. E., Newbolt, W. H., Gatewood, R. D., Foster, M. R., French, N. R., & Smith, D. E. (2000). Assessment center exercise factors represent cross-situational specificity, not method bias. *Human Performance, 13*, 323-353.
- 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*, 1202-1222.
- Lievens, F., & Goemaere, H. (1999). A different look at assessment centers: Views of assessment center users. *International Journal of Selection and Assessment, 7*(4), 215-219.

- Lowry, P. E. (1996). A survey of the assessment center process in the public sector. *Public Personnel Management, 25*, 307-321.
- Motowidlo, S. J., & Beier, M. E. (2010). Differentiating specific job knowledge from implicit trait policies in procedural knowledge measured by a situational judgment test. *Journal of Applied Psychology, 95*, 321-333. doi: 10.1037/a0017975
- 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*, 114-133. doi: 10.1037/a0030887
- Russell, C. J., & Domm, D. R. (1995). 2 Field-Tests of an Explanation of Assessment-Center Validity. *Journal of Occupational and Organizational Psychology, 68*, 25-47.
- Sebastian, R. A., Ramos, M. M., Stumbo, S., McGrath, J., & Fairbrother, G. (2014). Measuring Youth Health Engagement: Development of the Youth Engagement With Health Services Survey. *Journal Of Adolescent Health, 55*, 334-340. doi: 10.1016/j.jadohealth.2014.02.008
- Shore, T. H., Shore, L. M., & Thornton, G. C. (1992). Construct-Validity of Self Evaluations and Peer Evaluations of Performance Dimensions in an Assessment-Center. *Journal of Applied Psychology, 77*, 42-54.
- Spychalski, A. C., Quiñones, M. A., Gaugler, B. B., & Pohley, K. (1997). A survey of assessment center practices in organizations in the United States. *Personnel Psychology, 50*, 71-90.
- Thornton, G. C., & Krause, D. E. (2009). Selection versus development assessment centers: an international survey of design, execution, and evaluation. *International Journal of Human Resource Management, 20*, 478-498. doi: 10.1080/09585190802673536

Woehr, D. J., & Arthur, W. (2003). The construct-related validity of assessment center ratings: A review and meta-analysis of the role of methodological factors. *Journal of Management*, 29, 231-258. doi: Pii s0149-2063(02)00216-7

10.1016/s0149-2063(02)00216-7

Woehrmann, A. M., Deller, J., & Wang, M. (2014). A mixed-method approach to post-retirement career planning. *Journal of Vocational Behavior*, 84, 307-317. doi: 10.1016/j.jvb.2014.02.003

DRAFT

Table 1

The Perceived Nature and Incidence of Assessment Center Problems: Designers

Cat	Observed Problem	M
I	Assessors being asked to work extremely long hours	38
I	Assessors not being given the opportunity to practice the evaluation of exercises before the assessment centre went 'live'	25
I	Assessors asked to be involved in the assessment centre at the last minute, when they are were not properly prepared	23

Note. Cat = category; I = Aspects of the way the AC is implemented and run after it has been designed;

M = average proportions across all designers; Twenty-seven other problems, and their frequencies, are available from the first author on request.

DRAFT

Table 2

The Perceived Nature and Incidence of Pre-Consensus Meeting Problems: Assessors

Cat	Observed Problem	QO+
I	The order in which candidates take part in exercises not being the same for all	66
D	Candidates being assessed less than two times on a particular competence over the course of the assessment centre	53
I	Assessors being given insufficient time to evaluate and score candidates	49

Note. Cat = category; I = Aspects of the way the AC is implemented and run after it has been designed; D = AC design failings; A = assessor-related biases and errors, and poor behaviour by assessors; QO+ = percentage of assessors responses which were “quite often”, “very often” or “always or almost always”; Sixty other problems, and their frequencies, are available from the first author on request.

DRAFT

Table 3

*The Perceived Nature and Incidence of Consensus Meeting (i.e. 'Wash-up')
Problems: Assessors*

Cat	Observed Problem	QO+
A	Wash-ups which are rushed due to lack of time	51
A	More time spent discussing candidates at the beginning of the wash up than towards the end	42
A	Wash-ups in which there is too much unnecessary and unproductive discussion	40

Note. Cat = category; A = assessor-related biases and errors, and poor behaviour by assessors; QO+ = percentage of assessors responses which were "quite often", "very often" or "always or almost always".

Twenty-nine other problems, and their frequencies, are available from the first author on request.

DRAFT

Table 4

The Perceived Nature and Incidence of Assessment Center Problems: Candidates

Cat	Observed Problem	M
I	No clear guidance provided on how assessments will be marked	55
C	Not being given sufficient feedback on your performance after an assessment	51
D	Assessments which seemed poorly designed	34

Note. Cat = category; I = Aspects of the way the AC is implemented and run after it has been designed; D = AC design failings; C= Candidate consideration (lack of); O = Other; M = average percentage across all candidates. Thirty-nine other problems, and their frequencies, are available from the first author on request.

DRAFT

Table 5

The Frequency and Percentage of AC Problems Observed by Designers, Assessors and Candidates

Type of Problem	Designers		Assessors		Candidates	
	N	%	N	%	N	%
AC Design	16	53	10	11	10	24
AC Implementation	8	26	15	16	7	17
Assessor biases, errors, and poor behaviour	0	0	59	64	3	7
Training	1	3	4	4	2	5
Candidate consideration (lack of)	0	0	0	0	15	37
Other	5	17	4	4	4	10

Appendix 1

Instructions for Assessors

The following questions are concerned with your experience of being an assessor in an assessment center.

In each question you are asked how many times you have directly observed something.

If you are not sure, please give your best estimate.

Instructions for Candidates

Listed below are various events which you may, or may not, have experienced as a candidate in an assessment center.

Thinking of all the assessment centers you have attended as a candidate, in how many of these centers have you directly observed each of the following events?

Instructions for Designers

Listed below are various situations you may have experienced when designing assessment centers.

Thinking of all the assessment centers in which you have designed, or helped to design, in how many of these centers have you directly observed each of the following events?