Singapore Management University
Institutional Knowledge at Singapore Management University

12-1999

A different look at assessment centers: Views of assessment center users

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

Hans GOEMAERE
University of Ghent
DOI: https://doi.org/10.1111/1468-2389.00122

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

Part of the Industrial and Organizational Psychology Commons, and the Organizational Behavior and Theory Commons

Citation

PROFESSIONAL FORUM
A Different Look at Assessment Centers: Views of Assessment Center Users

Filip Lievens and Hans Goemaere*

This study aims to shed light on possible problems of assessment center users and designers when developing and implementing assessment centers. Semi-structured interviews with a representative sample of assessment center users in Flanders revealed that, besides a large variability in assessment center practice, practitioners experience problems with dimension selection and definition, exercise design, line/staff managers as assessors, distinguishing between observation and evaluation, and with the content of assessor training programs. Solutions for these problems are suggested.

Introduction

In the last 15 years assessment centers have become widespread in Western Europe, Northern America, and Australia (Newell and Shackleton, 1994). Although the make-up of assessment centers differs considerably from center to center, some common guidelines exist on how to develop assessment centers. More specifically, according to Thornton and Byham (1982) practitioners should first make decisions regarding the purpose of the assessment center (e.g., selection, development, etc.), the selection of the participants (e.g., self-nomination, supervisor nomination, pre-screening tests, etc.), and the target jobs to be included. Next, a set of dimensions thought to be necessary for effective job performance should be generated and a set of exercises to allow measurement of these dimensions should be constructed, resulting in an exercise by dimension matrix. The next stage should involve the selection and training of assessors. After these 'construction' stages attention should be paid to the operation of the assessment center. Assessors are expected to observe and record assessee behavior, classify these behaviors in dimensional terms, and provide dimensional ratings. Afterwards the assessors should gather to assign an overall assessment rating to each candidate (e.g., select, reject, promote, develop, etc.) by discussing behavioral observations and discrepancies in ratings. Finally, detailed behavioral feedback together with recommended actions should be provided to candidates and management.

To date, surprisingly little is known about which problems arise for assessment center users when implementing these stages of construction and operation (see Van Dam, Altink and Kok 1992, for an exception). Therefore, this study sought to examine the problems practitioners are facing in the design and implementation of assessment centers. In addition, it looked at whether these problems differ for organizations just starting to use assessment centers (less than three years), and for organizations that developed the assessment center without outside consultants.

Method

Sample

A representative sample of 176 large (more than 200 employees) organizations in Belgium (Flemish part) were telephoned and asked whether they used assessment centers for selection and/or development. All organizations contacted were private sector organizations. Forty-three (about 25%) of these organizations were using assessment centers. Members of the personnel department of 23 of these organizations agreed to be interviewed.

Of these 23 organizations, eight had developed their respective assessment centers without outside consultants. The remaining 15 relied on a consultancy agency to develop and implement the assessment center. Assessment centers were primarily developed for use within managerial and sales functions. Most assessment centers comprised a one-day programs with six participants.

Interview

The interview began with some general questions about the assessment center. Next,

* This article is based on Hans Goemaere's master's thesis, which was conducted at the University of Ghent. Filip Lievens is research assistant of the Fund for Scientific Research – Flanders (Belgium) (F.W.O.) Address for correspondence: Filip Lievens, Department of Personnel Management and Work and Organizational Psychology, University of Ghent, Henri Dunantlaan 2, 9000 Ghent, Belgium. E-mail: filip.lievens@rug.ac.be

for each assessment center design and operation stage (see model of Thornton and Byham, 1982) interviewees were asked to describe the process adopted and the problems experienced. On average, an interview lasted for about two hours. All interviews were audiotaped. Afterwards the interviews were transcribed and content analyzed by two industrial and organizational psychology students.

Results

Problems with Dimension Selection and Definition

Most large organizations in Flanders operated their assessment centers on the basis of a standard list of dimensions. The interviews revealed that three strategies were generally used to develop such lists. Some organizations simply adopted the dimension list of outside consultants. Others, gathering information from supervisors and job incumbents, conducted a job analysis of several target jobs to develop their own set of dimensions. A third group of organizations translated a set of 'strategic' dimensions determined by top management to the assessment center. Whatever strategy was used to compile the dimension list, the following 10 dimensions were most frequently used in large Flemish organizations: leadership, communicative skills, planning and organization, team orientation, social skills, problem analysis, initiative, creativity, tolerance for stress, and customer orientation.

Besides the fact that it was striking that only one third of the organizations conducted a thorough job analysis prior to the development of the assessment center, another noteworthy result emerging from the interviews pertained to the definition of the assessment center dimensions. In fact, practitioners were divided as to how specifically these dimensions should be defined. Some assessment center users preferred very elaborate definitions, whereby each dimension was defined and relevant behavioral observations were listed. However, other assessment center users criticized this fine-grained approach for not being generalizable to other exercises, jobs, and centers. Therefore, they only employed broad-grained dimensions and definitions applicable to various jobs and exercises. Then, it was left to the discretion of assessors to interpret the dimensions from exercise to exercise.

Assessment center users also mentioned problems with translating the dimensions to the exercises. Essentially, they reported that some assessment center dimensions (e.g., organizational sensitivity) were not really measurable as these dimensions usually generated few observable behaviors in the exercises. Related to this, most practitioners considered the attempt to measure each dimension in each exercise to be a serious pitfall in assessment center design. Similar thoughts were expressed with regard to the measurement of too many (i.e., more than 5) dimensions in a single exercise.

Problems with Exercise Selection and Development

The most popular simulation exercises were (in descending order) in-baskets, leaderless group discussions, role-plays (with irate customer or with problem subordinate), presentations, and case-analyses. The main problems here were related to exercise design. More specifically, the development costs and the lack of expertise to develop new exercises topped the comments received. This explains why Flemish assessment center users generally preferred generic exercises to bespoke exercises.

Problems with Assessor Selection and Training

It was found that in 50% of large Flemish organizations assessors exclusively comprised line/staff managers. In another 25% of organizations the assessor pool was composed of both line/staff managers and outside consultants. The remaining 25% of organizations completely relied upon outside consultants. The main argument for using line/staff managers as assessors was that line/staff managers evaluate assessees in terms of a broader frame-of-reference (i.e., organizational culture, future work environment, etc.). Yet it was noticeable that assessment center users also spotted problems with respect to the use of line/staff managers as assessors. For example, line/staff manager assessors tended to give rather global 'good/bad guy' judgments instead of distinct dimensional evaluations. When the assessment center was organized for promotion purposes, line/staff manager assessors were said to focus too much on the person (i.e., previous accomplishments, etc.) instead of the behavior observed.

One of the most striking results from the interviews related to the content of assessor training programs. Only half of the organizations provided a systematic training program. A possible explanation for this finding is that many of the organizations not investing in assessor training also completely relied on specialized consultancy agencies to develop the entire assessment center. Yet, time and cost considerations were also given as important reasons for not implementing systematic assessor training programs. When a training program did exist, the following components

were typically included: (a) introduction to the assessment center procedure (b), explanation of the dimensions and exercises used, (c) practice in the observation, evaluation, and integration processes.

According to Flemish assessment center designers the conduct and content of assessor training programs was not that straightforward. They posited that it was often very difficult for assessors to learn how to strictly separate observation and evaluation processes. Despite the plethora of practical exercises, the majority of trainees often confounded these two phases once they were dealing with real assessees. Another flaw in current training programs as noted by assessment center users related to the lack of refresher courses for assessors (In only one organization did refresher training actually exist). Nonetheless, Flemish assessment center designers considered such refresher courses to be crucial to update assessor proficiency, especially for managers only intermittently serving as assessors.

Problems with Observation and Rating Processes

Problems mentioned here paralleled the issues noticed in the context of assessor selection and training. For example, interviewees stated that assessors (especially line/staff managers) often made no distinction between observation and evaluation processes and gave global judgments. Another drawback was that prospective assessors typically recorded either too few or too many observations. Further, Flemish assessment center practitioners also noted that many of these problems arose because of planning or cost constraints. For instance, the ratio of assessees to assessors was often too large or assessors were required to evaluate too many ill-defined dimensions. Note that only half of the organizations were using behavioral checklists. Similarly, the interviews revealed that the videotaping of assessees has not really caught on in Flemish assessment center practice.

Besides these observation and evaluation issues, many practitioners reported huge efforts with respect to the concrete planning and operation of the assessment center. These efforts included, among others, the availability of a location, the availability of a pool of line/staff manager assessors, the rotation scheme of assessors across exercises, and the rotation scheme of assessees across exercises. A related problem was that the actual assessment center often did not run according to planning and timing. In order to tackle these operational issues, many practitioners were investing in information technology (e.g., PC in-basket, etc.).

Issues Related to the Overall Assessment Rating

In virtually all organizations the overall assessment rating and final dimension ratings were reached through a discussion between participating assessors. In only two organizations was a statistical decision rule used as the data integration method. Contrary to research results (e.g., Pynes and Bernardin 1992), practitioners assigned an added value to this time-consuming 'wash-up' discussion compared to a statistical decision rule. This added value was primarily related to the higher acceptability of the decisions made. Apparently, organizations considered this judgmentally-based data integration method, which lasted on average two and a half hours (per assessment center of six candidates), to be more important than quick decisions made by a mechanically-based decision rule.

Problems with Feedback to Candidates and Management

Practitioners did not describe specific problems regarding the provision of feedback to candidates and management. Yet the interviews revealed a lack of a systematic evaluation of the whole assessment center procedure. Although informal surveys of assessee reactions were not uncommon, a large-scale evaluation of the assessment center program in terms of reliability and validity was more the exception than the rule.

Differences among Organizations in Problems Experienced

First, we explored whether these problems differed for organizations with little assessment center experience (i.e., less than three years) compared to organizations with a lot of assessment center experience (i.e., more than three years). One of the specific problems noted by organizations just starting with assessment centers was that the assessment center approach had not yet been established in the organization. Consequently, personnel representatives of these organizations found it harder to organize assessor training and to gather line/staff managers for actual assessment centers. In addition, organizations just starting with assessment centers perceived the whole assessment center process as more time consuming than organizations with more assessment center experience.

Secondly, we also investigated whether organizations that developed the assessment center on their own experienced other problems which organizations using outside consultants did not. Our main finding was that, on the

whole, there existed a good relationship between organizations and their respective consultants. Organizations developing and conducting the assessment center themselves experienced somewhat more problems. For instance, as heretofore mentioned, these organizations were less inclined to invest in thorough assessor training programs. Furthermore, we found that the timing of assessment center exercises was a frequent stumbling block for these organizations.

Discussion

Whereas many previous studies focused on assessment center reliability and validity, this study aimed to shed light on possible problems of assessment center users and designers when developing and implementing assessment centers (for a similar study see Van Dam *et al.* 1992). This issue is related to what is referred to as the *practicality* of a selection instrument. Latham (1989) defines practicality as the ease or likelihood with which a group of users perceive that their objectives are attained through a particular selection instrument.

Besides the large variability in Flemish assessment center practice, the main problems occurred with dimension selection and definition, exercise design, using line/staff managers as assessors, distinction between observation and evaluation, and content of assessor training programs. These practical problems, which to some extent parallel those noted by Van Dam *et al.* (1992), challenge both researchers and practitioners to look for solutions. For instance, researchers have suggested that behavioral checklists may be helpful to side-step the aforementioned observation and rating problems (Reilly, Henry and Smither 1990). When assessors use behavioral checklists, they are not required to categorize behavior. Instead, they are able to concentrate their efforts on observation of relevant behaviors. Although some research suggests that such behavior checklists result in sharper dimension differentiation (Reilly *et al.* 1990), other studies found more equivocal results (see Lievens 1998, for a review). If checklists are used, prior research has demonstrated that behaviors should be ordered in naturally occurring clusters (Binning, Adorno and Kroeck, 1997) and that only the key behaviors should be listed (Hauenstein, 1994). With respect to the latter, Reilly *et al.* (1990) determined empirically that the optimal number of statements per dimension varies between six and twelve.

The concerns of Flemish assessment center users regarding the use of line/staff managers as assessors are supported by recent research. Sagie and Magnezy (1997) found that managers discriminated less between the dimensions than psychologists. Hence, we suggest that psychologists play a key role in assessor teams. For example, they may serve as coach of managerial assessors or as chair of the discussion session. Another possibility is to screen managers prior to serving as assessors. In this manner idiosyncratic assessors are identified in advance (see Hauenstein and Foti 1989, for a concrete example of this screening approach).

With respect to the problems inherent in assessor training, one of the striking findings was that one half of large Flemish organizations did not invest in a thorough assessor training program. This was partially due to the fact that many of these organizations completely relied on specialized consultancy agencies to develop the center. Yet, it is clear that using relatively untrained assessors detracts from the quality of the whole assessment center procedure. When time and cost constraints make it impossible for an organization to invest in assessor training, we suggest that such organizations install at least a system of assessor mentorship. This means that novice assessors play a secondary role in the assessment process and gradually learn 'the tricks of the trade' from experienced (psychologist) assessors and/or outside consultants.

Besides these recommendations suggested by assessment center researchers, other guidelines have been put forward by practitioners. For instance, a systematic model of exercise design has recently been developed, based on best practice (Ahmed, Payne and Whiddett 1997). More specifically, this model lists and specifies the important steps (from start to finish), criteria, and pitfalls in exercise development. Ahmed *et al.* (1997) note that this model enables assessment center users to develop an exercise in one morning. Given these advantages this systematic model may be very fruitful in answering the exercise construction concerns reported above.

References

Ahmed, Y., Payne, T. and Whiddett, S. (1997) A process for assessment exercise design: a model of best practice. *International Journal of Selection and Assessment*, **5**, 62–68.

Binning, J.F., Adorno, A.J. and Kroeck, K.G. (1997) *Validity of behavior checklist and assessor judgmental ratings in an operational assesment center.* Paper presented at the Annual Conference of the Society for Industrial and Organizational Psychology, St Louis, MO, April.

Hauenstein, P.C. (1994) *A key behavior approach for improving the utility of developmental assessment centers.* Paper presented at the Annual Conference of the Society for Industrial and Organizational

Psychology, Nashville, TN, April.

Hauenstein, N.M.A. and Foti, R.J. (1989) From laboratory to practice: neglected issues in implementing frame-of-reference rater training. *Personnel Psychology*, **42**, 359–378.

Latham, G.P. (1989) The reliability, validity, and practicality of the situational interview. In R.W. Eder and G.R. Ferris (eds.) *The Employment Interview: Theory, Research, and Practice* (pp. 169–182). Thousands Oaks, CA, Sage.

Lievens, F. (1998) Factors which improve the construct validity of assessment centers: A review. *International Journal of Selection and Assessment*, **6**, 141–152.

Newell, S. and Shackleton, V. (1994) Guest Editorial: International differences in selection methods. *International Journal of Selection and Assessment*, **2**, 71–73.

Pynes, J. and Bernardin, H.J. (1992) Mechanical vs. consensus-derived assessment center ratings: A comparison of job performance validities. *Public Personnel Management*, **21**, 17–28.

Reilly, R.R., Henry, S. and Smither, J.W. (1990) An examination of the effects of using behavior checklists on the construct validity of assessment center dimensions. *Personnel Psychology*, **43**, 71–84.

Sagie, A. and Magnezy, R. (1997) Assessor type, number of distinguishable categories, and assessment centre construct validity. *Journal of Occupational and Organizational Psychology*, **70**, 103–108.

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

Van Dam, K., Altink, W.M.M. and Kok, B. (1992) Assessment center practice: A summary of problems encountered [In Dutch]. *De Psycholoog*, **7**, 509–514.

Woehr, D.J. and Huffcutt, A.I. (1994) Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organisational Psychology*, **67**, 189–205.