

Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection Lee Kong Chian School Of
Business

Lee Kong Chian School of Business

1-1999

Development of a simulated assessment center

Filip LIEVENS

Singapore Management University, filiplievens@smu.edu.sg

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



Part of the [Human Resources Management Commons](#), and the [Organizational Behavior and Theory Commons](#)

Citation

LIEVENS, Filip. Development of a simulated assessment center. (1999). *European Journal of Psychological Assessment*. 15, (2), 117-126.

Available at: https://ink.library.smu.edu.sg/lkcsb_research/5703

This Journal Article is brought to you for free and open access by the Lee Kong Chian School of Business at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection Lee Kong Chian School Of Business by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylds@smu.edu.sg.

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/232430397>

Development of a Simulated Assessment Center

Article in *European Journal of Psychological Assessment* · May 1999

DOI: 10.1027//1015-5759.15.2.117

CITATIONS

5

READS

108

1 author:



Filip Lievens

Singapore Management University

279 PUBLICATIONS 9,262 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Modular approach to selection [View project](#)



Quantification of human mimicry to improve understanding of applicant-interviewer dynamics. [View project](#)

DEVELOPMENT OF A SIMULATION OF AN ASSESSMENT CENTER

Filip Lievens

University of Ghent

This study was supported by a grant of the Fund for Scientific Research-Flanders (F.W.O.). I acknowledge Marise Born, Pol Coetsier, Wilfried De Corte, Charles Lance, George Thornton III, and two anonymous reviewers for their helpful suggestions on an earlier version of this paper. I am also especially grateful to Herlinde Pieters, Luc Drieghe, and Lieven Verstraete for their assistance in conducting the assessment center simulation sessions. Parts of this article were presented at the 4th European Conference on Psychological Assessment, Lisbon, 1997.

Correspondence concerning this article should be addressed to Filip Lievens, Department of Personnel Management and Work and Organizational Psychology, University of Ghent, Henri Dunantlaan 2, 9000 Ghent, Belgium. Electronic mail may be sent via Internet to [filip.lievens@rug.ac.be].

Lievens, F. (1999). Development of a simulated assessment center, *European Journal of Psychological Assessment*, 15, 117-126.

Summary

In this study a simulation of an assessment center is developed consisting of videotaped performances of four candidates in three exercises: sales presentation, role-play, and group discussion. To develop the simulation, candidate profiles are constructed which vary along three dimensions: problem analysis and solving, interpersonal sensitivity, and planning and organizing. Dimension-specific behaviors are adapted from 173 critical incidents provided by 20 experienced assessors. Resulting from this, scripts are written and enacted by actors. To validate the simulation, experienced assessors evaluate all candidates under optimal conditions. High inter-rater agreement among these experts (intraclass correlation = .9) and a high correlation ($r = .93$) between expert scores and intended scores are found. Finally, the simulation is piloted and calibrated on 16 managers and 28 industrial psychology students. Practical applications of the simulation involve assessor training and certification. Researchers may use it to examine which factors affect the accuracy of assessor judgments.

**KEYWORDS: ASSESSMENT CENTERS, TRAINING, SIMULATION,
RATING PROCESSES**

DEVELOPMENT OF A SIMULATION OF AN ASSESSMENT CENTER

Introduction

In the last forty years assessment centers (ACs) have become a popular approach for managerial selection and assessment (Spsychalski, Quinones, Gaugler, & Pohley, 1997). Most ACs consist of situational tests that represent tasks frequently performed by managers. The situational tests (e.g., role-plays, group discussions) assess how well candidates perform on a number of job-related dimensions.

Examples of these dimensions are planning, problem solving, and sensitivity.

The raters, also known as assessors, are key factors of ACs. For instance, assessors observe, record and rate the various candidates. Next, assessors meet to integrate the information into an overall judgment. Assessors also write the final reports and give feedback to candidates. Both practitioners and researchers acknowledge the demanding and central role of assessors in the AC process. Hence, from a research perspective a plea is often made for studies focusing on the assessor (Bartels & Doverspike, 1997; Guion, 1987; Klimoski, 1993; Lievens, 1997; Zedeck, 1986). Practitioners on the other hand are concerned that, prior to serving as assessors in ACs, individuals participate in a practice oriented assessor training (Byham, 1977; Task Force on Assessment Center Guidelines, 1989).

This study aims to meet both of these concerns. Our purpose was to develop a video-based simulation of an AC for use as stimulus material in both research and training. This article begins by detailing the practical and scientific need for such a simulation. Next, we outline the methodology adopted to simulate the assessor task and context. Finally, we present the final simulation and discuss its usability.

Practical Need for Simulating Assessor Environment

Because serving as an assessor in ACs is both such a complex and important task, assessors have to attend a training program (Task Force on Assessment Center Guidelines, 1989). In a seminal chapter Byham (1977) set a high standard for the content and conduct of an assessor training. As a minimum, the training should provide assessors with a thorough understanding of the essentials of ACs. More important, the training should help to master skills such as acquiring, evaluating, and communicating information about people. In this context Byham wrote:

There is no substitute for giving an assessor the opportunity to put all of the training elements together in an actual exercise. This practice is the mainstay of most assessor training programs. The assessors are given the opportunity to observe an exercise, using practice subject(s) or videotape. They record behavior as if they were assessing the individual(s) and then complete the observation forms. (p. 104)

Existing training programs, however, often use short written vignettes of candidate behavior to exercise the classification and evaluation skills of assessors. A drawback of the vignettes is that they fail to capture the liveliness of candidates. Another limitation of current programs is that assessors are seldom evaluated at the end of the training track (Spsychalski et al., 1997). Objective standards to benchmark assessor ratings are also seldom available.

These limitations in existing training programs are met by carefully developing videotapes of assessee performances. Such a video-based AC simulation is useful (a) to train assessors in observing and rating candidates, or (b) to evaluate assessors' rating proficiency. In both cases, assessors should receive feedback on how well their ratings correspond to expert ratings.

Methodological need for simulating Assessor Environment

The vast majority of earlier AC research was geared towards answering the question whether ACs worked (Howard, 1997). This large body of research demonstrated that ACs worked and were good predictors of a variety of criteria in the domain of managerial success (see Gaugler, Rosenthal, Thornton, & Bentson, 1987 for a meta-analytic review). Some areas, however, remain underinvestigated. For instance, little is known about how ACs exactly work (Klimoski, 1993). Hence, calls have been made to shift research challenges from "Do ACs work?" to "How do ACs work?". In other words, more process oriented studies focusing on the pivotal role of the assessor should be conducted (Bartels & Doverspike, 1997; Guion, 1987; Klimoski, 1993; Zedeck, 1986). These researchers have argued to investigate the following issues: Are assessor judgments of candidates accurate? Do assessors use all dimensions when rating? How do assessor characteristics impact on the quality of judgments made? Are assessor judgments subject to rating tendencies (halo, contrast effects, etc.)?

Research on assessor judgments and processes is typically investigated in laboratory settings. Then, the stimulus material consists of videotapes of hypothetical candidates which are rated by assessors (e.g., Gaugler & Rudolph, 1992; Ryan et al., 1995). These lab studies are often vulnerable to external validity and, hence, do not generalize to the 'real world'. For instance, prior to rating candidates, assessors are not given information about the organization which would hire the candidates (see Gaugler & Rudolph, 1992; Ryan et al., 1995).

To meet these external validity concerns, Funder (1987) advocated for faithfully reconstructing all of the important elements and sources of information, that actually are found in a particular real-life situation. Similarly, Sackett and Larson (1991) strongly argued to set up simulation experiments. A simulation experiment verses the participant into a high-fidelity reconstruction of a real-life

situation. Hence, a major thrust of this study was to carefully map the characteristic elements of assessors' task and environment prior to the development of the simulation.

Objectives of this Study

In sum, this study aims to develop a video-based simulation of an AC. To be useful as stimulus material in both research and training, the simulation should fulfill three requirements: external validity, internal validity, and feasibility.

1. To increase external validity, the AC simulation should be representative of the typical assessor task and context.
2. Regarding internal validity, the AC simulation should be developed according to the 'true score' paradigm (Borman, 1977). This implies that the 'true' performance levels of the videotaped candidates are known.
3. It should be feasible for assessors to complete the simulation without fatigue.

Method

Characteristics of Assessor Task and Context

The main task of assessors consists in observing and recording (non) verbal candidate behavior. Most assessors record ongoing behavior ('direct observation'). Nonetheless, assessors may also indirectly observe videotaped performances of candidates (Bray & Byham, 1991). After taking notes, assessors rate candidates on multiple job-related dimensions. The recorded and to a lesser extent the recalled behaviors serve as input for rating. Therefore, the rating process of ACs is rather stimulus-based than memory-based (Borman & Hallam, 1991). Both psychologists and line managers serve as assessors. A typical ratio of assesseees to assessors is 2

to 1. (Thornton, 1992). Besides observing and rating, assessors often engage in role-play activities (see Zedeck, 1986), discuss the ratings with other assessors, prepare final reports, and provide feedback.

Assessors do not evaluate candidates in a vacuum. There are a variety of contextual elements which may affect assessor ratings. (see Murphy & Cleveland, 1995). For instance, assessors evaluate candidates in job-related exercises. Typical for these situational exercises is their limited time span (15 to 45 minutes), and that a steady stream of assessee behavior is presented at a very fast rate (Gaugler & Thornton, 1989). To aid this complex task, assessors often use behavioral checklists (Reilly, Henry & Smither, 1990). Another contextual influence is the rating purpose. Assessors may evaluate candidates differently, depending on whether their ratings will serve a selection purpose (i.e., 'yes/no' decision) or a developmental purpose (i.e., identification of strengths and weaknesses). The influences of assessor-assessee acquaintance should be minimal, because assessors generally do not know the assessees (Schuler, Moser, & Funke, 1994) and because assessors rely more on information elicited first-hand (i.e., observational data) than to biodata or psychometric test scores (Anderson, Payne, Ferguson, & Smith, 1994). The motivational level (i.e., accountability) of assessors is high. Afterwards they are required to justify their ratings to fellow assessors, assessees, and organization (Gaugler & Rudolph, 1992). Finally, the organizational context should not be neglected. Both objective (e.g., type of business, location, level of decentralization, size of the workforce, etc.) and subjective (e.g., organizational culture) factors may influence assessor judgments.

Sampling of Characteristic Elements in Assessment Center Simulation

As mentioned above, the simulation should be representative of the assessor task and assessor context. Using documents of operational ACs, we reconstructed the various contextual elements. For instance, actual rating scales and AC exercises (i.e., sales presentation, role-play, and group discussion) are used. The job posting of a district sales manager is also based on a real job posting. Information about the organizational context is adapted from an annual report of an existing company.

To simulate the assessor task, we developed scripts which depicted the behavior of four candidates in the three exercises: Candidate profiles vary along three dimensions: problem analysis and solving, interpersonal sensitivity, and planning and organizing. To this end, we followed procedures outlined by Borman (1977) and McCauley et al. (1990). First, a representative pool of assessee behavior for each dimension in each of the three AC exercises was gathered. Twenty assessors (15 males; mean age = 36 years) were asked to provide behaviors that cause them to judge an assessee as being higher or lower. The assessors qualified as experts due to (a) their practical experience as assessors (mean assessor experience¹ = 6 years), (b) their theoretical knowledge of ACs, and (c) their familiarity with AC research. These experts generated a total of 765 (non) verbal assessee behaviors. Additionally, rating forms of five psychological consulting firms were scrutinized, resulting in 121 behaviors. After eliminating redundancies, the total list of 886 behaviors was reduced to 310 behaviors.

Second, ten other assessors (8 males; mean age = 37 years; mean assessor experience = 7 years) considered each behavior for a given exercise and marked the dimension to which that behavior belonged. A criterion of 80% agreement was used to select behaviors for each dimension within each exercise (Reilly et al., 1990). As shown in table 1, 173 behaviors survived the retranslation process. As an example, appendix 1 lists the 'retranslated' behaviors of the role-play.

Insert table 1 about here

Third, we determined the intended ‘true’ scores per candidate. These scores indicate whether a candidate performs well (= 5), moderately (= 3) or poorly (= 1) on dimensions. For example, the first column of table 2 shows that candidate 1 is highly sensitive to others (= 5) and pays hardly attention to planning and organization (= 1). His analysis and problem solving qualities are moderate (= 3). Across the three exercises each candidate has the same performance profile. Within exercises the candidate profiles are heterogeneous (i.e., a candidate did not score highly on all three dimensions).

Insert table 2 here

Fourth, we wrote scripts of the AC performance of the four candidates. The second and third column of table 2 illustrate how we developed the scripts. We selected per dimension five or six behaviors from the pool of incidents gathered. Then, these behaviors were built into the scripts. Care was taken to preserve realism and smoothness by nesting the critical behaviors among innocuous material. The scripts depicted the word-for-word dialogue for each performance. Nine scripts were written: four scripts of a candidate delivering a sales presentation, four scripts of the same four candidates talking to a disgruntled employee, and one script of a discussion between these assesseees. Two experienced assessors tested the scripts for realism and made adjustments. Appendix 2 presents the final version of a script.

Fifth, semi-professional actors were filmed delivering their scripted AC performances. Logically, the same actor enacted the same candidate across the three exercises. Prior to filming, actors were briefed about the target job and ACs.

Finally, the intended true scores (see step 3) were validated by comparing them to ‘expert scores’ (see Sulsky & Balzer, 1988). An expert score is “the rating

that would be expected from an unbiased, careful rater who completed the rating task under optimal conditions” (Murphy & Cleveland, 1995, p. 285). To this end, five experienced assessors (3 males; mean age = 30 years; mean assessor experience = 4 years) reached consensus over what constituted effective and ineffective performances on the various dimensions. Next, these experts observed each performance and recorded observations. They could view the tapes repeatedly and rewind them. They were also provided with the scripts. All experts independently rated each videotaped performance on a 5 point scale, with 1 poor indicating and 5 indicating excellent. Table 3 presents the scores estimated by the experts and the intended true scores. Interrater agreement among the expert ratings equaled .9 (intraclass correlation 2.1, Shrout & Fleiss, 1979). After discussing rating discrepancies, the experts agreed on a set of final expert scores (see the last column of table 3). These expert scores correlated highly with the intended true scores ($r = .93$), demonstrating that the videotaped performances carefully reflect the intended scores.

Insert table 3 about here

Results

Structure of Final AC simulation

The final film used in the AC simulation consists of four parts. In the first introductory part participants learn that they will serve as assessors in an AC simulation. They will watch four candidates applying for the job of ‘district sales manager’ in the organization ‘Plafox’. Passport photos of the candidates are shown. Assessors hear that the district sales manager will be responsible for developing and maintaining business partnerships within a specified geographical area. He will also manage a team of four junior sales representatives. Candidates

need a BS/BA degree in Engineering, five years of experience in investment goods sales, and preferably some experience as a leader. Assessors learn that Plafox is a leading manufacturer of partitions and false ceilings, with over 1,100 employees in 30 agencies worldwide. Pictures of the products made are displayed. Plafox promotes itself as a young, dynamic and customer-centered company. Next, assessors are familiarized with the AC dimensions, exercises and 5-point graphical rating scales. Assessors receive a list of 12 dimensions which a job analysis identified to be crucial for effective district sales managers. Assessors are informed that three performance dimensions are of paramount importance: analysis and problem solving, interpersonal sensitivity, and planning and organizing. Candidates will perform in three situational exercises. The introductory film ends by stating that assessors have to independently rate the candidates. After all videotapes, they will discuss their ratings with colleagues.

The second part of the film shows how each candidate delivers a sales presentation (average length = 6 minutes). In this exercise each candidate has to present an in-depth analysis of the buyer's needs and thoroughly argue which of three software systems is most appropriate. The buyer is represented by a panel of decision makers. Afterwards, the panel asks questions to challenge the candidate.

The third part displays each candidate in a role-play (average length = 5 minutes) with David. Recently, several clients were not pleased about David. The candidates are expected to dig into the reasons for the complaints and to suggest solutions. The candidates do not know that David is disappointed for not being admitted to a management course. Appendix 2 presents the dialogue between candidate 1 and David (see table 2 for true scores and behavioral incidents).

Finally, the film shows the four candidates in a business meeting (14 minutes). Each candidate manages a division of an organization. The managers

meet to divide next year's budget. Prior to the meeting, each of them was assigned a series of projects. The candidates do not know the projects of one another. As the budget is restricted, the managers have to reach consensus on which projects to select.

All together, these nine videotaped performances run 55 minutes. No technical knowledge is required from assessors to rate the candidates in the exercises.

Calibration of simulation

The simulation was piloted on 28 industrial psychology students (9 men, mean age = 21.4 years, SD = 0.6 years) and 16 managers (8 men, mean age = 36.5 years, SD = 6.9 years). With regard to the students, the simulation was an optional course on personnel selection (all students had nominated themselves to participate). The managers, which came from different organizations, had subscribed to attend a 3-day program on personnel assessment. The simulation was a part of this program.

Besides pilot testing the simulation, we were interested in the participant's accuracy in rating candidates. Per participant a differential accuracy index (Cronbach, 1955) was computed, for each of the three exercises rated. Differential accuracy indicates how accurately candidates are rank ordered on the various dimensions. To this end, each participant's ratings are contrasted to the true scores (for exact formulas see Cronbach, 1955). Lower values on differential accuracy indicate higher accuracy, whereas higher values reflect lower accuracy. Because the managers had already experience in rating subordinates, we expected them to be more accurate than the students. A t-test showed significant differences in differential accuracy between the two groups for rating the presentation, $t(42) = 3.08$, $p < .01$, role-play, $t(41) = 3.01$, $p < .01$, and discussion, $t(39) = 2.70$. $p < .05$.

The distribution of differential accuracy indices (not reported here for the sake of brevity) may be used as a norm table to benchmark ratings of future assessors.

Discussion

This study aimed to develop a video-based AC simulation for use as stimulus material in assessor research and training. From a research perspective, this simulation opens new perspectives in studying ACs. Whereas previous research was mainly done on a higher level of aggregation (Guion, 1987), the simulation could promote an individual level analysis approach. Because true scores were built into the simulation, it is particularly interesting for studying the accuracy of individual assessors. To date, the accuracy of assessor judgments has rarely been a dependent variable (e.g., Gaugler & Thornton, 1989; Ryan et al., 1995). In addition, research on the factors which affect assessor accuracy is needed. Process oriented AC studies also might give researchers some clues regarding the lack of convergent and discriminant validity of ACs (Klimoski, 1993).

For practical purposes the video-based simulation may serve as an experiential exercise to improve observation and rating skills. For instance, after a lecture about AC basics, participants may evaluate the videotaped performances. Next, assessors may meet in teams to share observations, discuss ratings, and write candidate reports. This training approach is a springboard for rating real people, because it focuses on practice and places participants in a simulated assessor environment. Practitioners may also use the AC simulation for selection or certification. Then, the simulation is an objective test to measure the rating proficiency of assessors (Task Force on Assessment Center Guidelines, 1989). Irrespective of which option selected, it is crucial that the participating assessors receive objective feedback about their rating performance. To this end, ratings of each assessor are compared to the true ratings and to the related behavioral

incidents (e.g., table 2). These true ratings and behavioral rationales enable trainers to give participants feedback about their differential accuracy (e.g., for which exercise were ratings inaccurate) and their behavioral accuracy (e.g., which behaviors remained unnoticed). Despite these advantages, a limitation of the simulation is that videotaped instead of 'live' candidates are used. Nonetheless, videotaping assesses has become widespread in ACs (Bray & Byham, 1991). Moreover, Ryan et al. (1995) found no significant differences in accuracy between direct and indirect observation of assessees.

In sum, in this study both rating task and context of assessors were carefully reconstructed and simulated. The video-based AC simulation enables researchers to investigate process oriented AC issues. Practitioners may find this simulation helpful for assessor training and certification.

Footnotes

¹ All assessors were well versed in AC practice because they worked in psychological consulting firms specialized in ACs. There was a high similarity between the exercises of the simulation and the exercises in their ACs. Mean ratings on 9-point scales (1 = no similarity, 9 = high similarity) for the presentation were 7.7 (SD = 1.4), for the role-play 7.5 (SD = 1.8), and for the discussion 7.3 (SD = 2.2).

² The videotaped candidate performances are available for research and educational purposes only from the author.

References

Anderson, N., Payne, T., Ferguson, E. & Smith, T. (1994). Assessor decision making, information processing and assessor decision strategies in a British assessment centre. Personnel Review, *23*, 52-62.

Bartels, L.K. & Doverspike, D. (1997). Assessing the assessor, the relationship of assessor personality to leniency in assessment center ratings, Journal of Social Behavior and Personality, *12*(5), 179-190.

Borman, W.C. (1977). Consistency of rating accuracy and rater errors in the judgment of human performance. Organizational Behavior and Human Performance, *20*, 238-252.

Borman, W.C. & Hallam, G.L. (1991). Observation accuracy for assessors of work-sample performance: Consistency across task and individual-differences correlates. Journal of Applied Psychology, *76*, 11-18.

Bray, D.W. & Byham, W.C. (1991). Assessment centers and their derivatives. The Journal of Continuing Higher Education, *39*, 8-11.

Byham, W.C. (1977). Assessor selection and training. In J.L. Moses & W.C. Byham (Eds.), Applying The Assessment Center Method. (pp. 89-125). New York: Pergamon Press.

Cronbach, L.J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity". Psychological Bulletin, *52*, 177-193.

Funder, D.C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. Psychological Bulletin, *101*, 75-90.

Gaugler, B.B. & Rudolph, A.S. (1992). The influence of assessee performance variation on assessors' judgments. Personnel Psychology, *45*, 77-98.

Gaugler, B.B. & Thornton, G.C. (1989). Number of assessment center dimensions as a determinant of assessor accuracy. Journal of Applied Psychology, 74, 611-618.

Gaugler, B.B., Rosenthal, D.B., Thornton, G.C. & Bentson, C. (1987). Meta-analysis of assessment center validity. Journal of Applied Psychology, 72, 493-511.

Guion, R.M. (1987). Changing views for personnel selection research. Personnel Psychology, 40, 199-213.

Howard, A. (1997). A reassessment of assessment centers, challenges for the 21st century. Journal of Social Behavior and Personality, 12(5), 13-52.

Klimoski, R. J., (1993). Predictor constructs and their measurement. In N. Schmitt & W.C. Borman (Eds.), Personnel Selection in Organizations (pp. 99-134). San Francisco: Jossey-Bass.

Lievens, F. (1997). Can we improve the construct validity of assessment centres? Paper presented at the 25th International Congress of Assessment Center Methods, London, UK.

McCauley, D.P., Jago, I.A., Gore, B., Lance, C.E., Quarles, F.K., Sledge, L., Pate, J.L., Logan, A.L. & Guest, F.D. (1990). The development of more ecologically valid videotapes for use in performance appraisal research. Paper presented at the annual meeting of the Southeastern Psychological Association, Atlanta, GA.

Murphy, K.R. & Cleveland, J.N. (1995). Understanding performance appraisal. Thousands Oaks: Sage.

Reilly, R.R., Henry, S. & 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.

Ryan, A.M., Daum, D., Bauman, T., Grisez, M., Mattimore, K., Nalodka, T. & McCormick, S. (1995). Direct, indirect, and controlled observation and rating accuracy. Journal of Applied Psychology, 80, 664-670.

Sackett, P.R. & Larson, J.R. (1991). Research strategies and tactics in industrial and organizational psychology. In M.D. Dunnette & L.M. Hough (Eds), Handbook of Industrial and Organizational Psychology (pp. 419-489). Palo Alto: Consulting Psychologists Press Inc.

Schuler, H. Moser, K. & Funke, U. (1994). The moderating effect of rater-ratee acquaintance on the validity of an assessment center. Paper presented at the 23th International Congress of Applied Psychology, Madrid, Spain.

Shrout, P.E. & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 86, 420-428.

Spychalski, A.C., Quinones, M.A., Gaugler, B.B & Pohley, K.A. (1997). A survey of assessment center practices in organizations in the United States, Personnel Psychology, 50, 71-90.

Sulsky, L.M. & Balzer, W.K. (1988). The meaning and measurement of performance rating accuracy: Some methodological concerns. Journal of Applied Psychology, 73, 497-506.

Task Force on Assessment Center Guidelines (1989). Guidelines and ethical considerations for assessment center operations. Public Personnel Management, 18, 457-470.

Thornton, G.C., III (1992). Assessment centers in Human Resource Management. Reading: Addison-Wesley Publishing Company.

Zedeck, S. (1986). A process analysis of the assessment center method. Research in Organizational Behavior, 8, 259-296

TABLE 1

Number of Behavioral Incidents by Dimension and Exercise

Dimension	Prior to retranslation phase			After retranslation phase		
	PRES ^a	ROLE ^b	DISC ^c	PRES	ROLE	DISC
Analysis and Problem Solving	28	47	42	16	23	22
Interpersonal Sensitivity	26	49	33	17	26	20
Planning and Organizing	18	37	30	12	18	19
Total	72	133	105	45	67	61

Note ^a PRES = Sales Presentation; ^b ROLE = Role-play with a disgruntled employee;

^c DISC = Group discussion

TABLE 2

Intended True Scores and Behavioral Incidents of Candidate 1 in Role-play

Intended true score on dimensions	Behavioral incident	Concrete example in script
<u>Moderate</u> Analysis and Problem Solving	<ul style="list-style-type: none"> • Asking the subordinate questions about his/her work situation and recent behavior. • Keeping on asking questions when vague or ambiguous answers are given. • Not finding the underlying cause(s) of problem. • Suggesting multiple solutions to problem. • <u>Not</u> considering the pros and cons of various options. 	<p>‘What has happened there?’; ‘What have you done recently?’</p> <p>‘Nothing at all ...?’</p> <p>‘In my opinion, when the deal was almost done, and surely when the client was ... Fyrens’; ‘Perhaps, we have found the bottleneck. If you are about to make the deal with Fyrens ...’</p> <p>‘You have to clear the sky ...resolve the problem as soon as possible’; ‘you should manage your time ... better?’</p>
<u>High</u> Interpersonal Sensitivity	<ul style="list-style-type: none"> • Using effective body language/facial expressions to show listening (nodding, etc.). • Asking questions regarding the subordinate’s feelings. • Showing consideration for the situation of subordinate. • Emphasizing also the strengths of subordinate. • Involving subordinate into problem solving process. • Asking for approval of the subordinate. 	<p>‘What’s wrong?’; ‘and therefore, you are a bit ...’</p> <p>‘I understand’; ‘I understand you are upset.’</p> <p>‘I am convinced you will succeed. In the past, you have proven this ...’; ‘I still find you a very good salesman’; ‘Well, success, David!’</p> <p>‘How do we resolve this together?’; ‘What do you suggest?’</p> <p>‘Do you follow me?’; ‘What do you think?’; ‘What do you mean?’; ‘Is everything clear ...?’</p>

<u>Low Planning and Organizing</u>	<ul style="list-style-type: none">• The specific goals of the interview are not formulated at the beginning.• A specific agenda for the meeting is lacking.• conducting the interview in a rather unstructured way.• Making <u>neither</u> concrete <u>nor</u> specific agreements • Concrete deadlines are not formulated.	<p><u>‘By the way, what kind of meeting is this?’</u> ‘Sorry, I had to make this clear to you, right from the beginning’</p> <p>‘I would like to discuss the course of things.’</p> <p>‘Euh, before I forget, David. Don’t you think you should ...?’</p> <p><u>‘What do I have to do exactly from now on?’</u> ‘...clear the sky’; ‘ resolve the problem as soon as possible ...’</p>
------------------------------------	---	--

Note. See appendix 2 for the complete script of this candidate’s role-play performance.

TABLE 3
Intended and Estimated True Scores of Videotaped Candidates

Dimension	True scores						
	Intended	Estimated by experts					
		Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	After discussion
	Candidate 1						
Analysis and Problem Solving (PRES) ^a	3	3	3	3	3	3	3
Interpersonal Sensitivity (PRES) ^a	5	5	5	5	5	4	5
Planning and Organizing (PRES) ^a	1	2	2	1	1	2	2
Analysis and Problem solving (ROLE) ^b	3	3	3	3	3	3	3
Interpersonal Sensitivity (ROLE) ^b	5	5	4	5	4	4	5
Planning and Organizing (ROLE) ^b	1	2	2	2	2	2	2
Analysis and Problem solving (DISC) ^c	3	3	3	4	3	3	3
Interpersonal Sensitivity (DISC) ^c	5	5	5	5	5	4	5
Planning and Organizing (DISC) ^c	1	2	2	2	2	2	2
	Candidate 2						
Analysis and Problem Solving (PRES)	1	2	3	2	2	4	2
Interpersonal Sensitivity (PRES)	3	3	2	2	3	3	2
Planning and Organizing (PRES)	5	5	5	5	5	5	5
Analysis and Problem solving (ROLE)	1	1	2	2	1	2	2
Interpersonal Sensitivity (ROLE)	3	2	2	2	2	2	2
Planning and Organizing (ROLE)	5	5	5	5	5	4	5
Analysis and Problem solving (DISC)	1	2	3	1	1	3	2
Interpersonal Sensitivity (DISC)	3	2	2	2	2	3	2
Planning and Organizing (DISC)	5	5	5	5	5	4	5

	Candidate 3						
Analysis and Problem Solving (PRES)	5	5	5	5	5	4	5
Interpersonal Sensitivity (PRES)	1	1	1	1	1	2	1
Planning and Organizing (PRES)	3	4	3	4	3	3	4
Analysis and Problem solving (ROLE)	5	5	4	5	5	4	5
Interpersonal Sensitivity (ROLE)	1	2	2	1	2	2	2
Planning and Organizing (ROLE)	3	4	3	4	3	3	4
Analysis and Problem solving (DISC)	5	5	4	5	5	4	5
Interpersonal Sensitivity (DISC)	1	1	1	1	1	2	1
Planning and Organizing (DISC)	3	4	4	4	4	4	4
	Candidate 4						
Analysis and Problem Solving (PRES)	1	2	2	2	2	2	2
Interpersonal Sensitivity (PRES)	1	1	1	1	1	1	1
Planning and Organizing (PRES)	1	2	2	1	2	2	2
Analysis and Problem solving (ROLE)	1	1	1	2	1	2	1
Interpersonal Sensitivity (ROLE)	1	1	1	1	1	1	1
Planning and Organizing (ROLE)	1	2	2	2	2	3	2
Analysis and Problem solving (DISC)	1	1	1	1	1	2	1
Interpersonal Sensitivity (DISC)	1	1	1	1	1	2	1
Planning and Organizing (DISC)	1	1	1	2	2	2	2
\underline{r} (intended and estimated true scores)	--	0,94	0,89	0,94	0,95	0,84	0,93
95% confidence interval for \underline{r} (lower bound)	--	0,89	0,79	0,88	0,89	0,71	0,87
95% confidence interval for \underline{r} (upper bound)	--	0,97	0,94	0,97	0,97	0,92	0,96

Note. Candidate ratings were made on 5-point scales, with 1 indicating a low level of the dimension and 5 indicating a high level of the dimension. ^a PRES = Presentation; ^b ROLE = Role-play, ^c DISC = Group discussion.

APPENDIX 1

Retranslated Behavioral Incidents for Three Dimensions in Role-Play

Analysis and Problem Solving

1. relying upon the available information and facts when making decisions (6)
2. recognizing when additional information is needed and looking for that information (2)
3. asking subordinate questions about his/her work situation and recent behavior (2)
4. exploring the motives and reasons of what has happened (through precise and specific questions) (13)
5. asking subordinate to evaluate his/her current (and former) effectiveness (4)
6. asking questions about how others (e.g., colleagues) react to his/her recent behavior (1)
7. asking questions about boundary conditions (1)
8. keeping on asking questions when vague or ambiguous answers are given (12)
9. making logical connections among different pieces of information (7)
10. approaching the problem from various angles (4)
11. being able to separate essentials from trivialities(4)
12. indicating what really matters (3)
13. taking also the relevant details into consideration (1)
14. finding the underlying cause(s) of problem (4)
15. indicating (the cause and) consequences of the subordinate's behavior (3)
16. considering critically the various proposals, options, and solutions (5)
17. considering the pros and cons of various options (2)
18. formulating suggestions in a hypothetical way (1)
19. being able to synthesize and to take decisive action (9)
20. trying to obtain a solution which is in favor of both parties (1)
21. suggesting multiple solutions to problem (1)
22. understanding the different components of a problem before developing solutions (1)
23. signaling that information is missing (1)

Interpersonal Sensitivity

1. eye-contact with subordinate (7)
2. using body language/facial expressions to show listening (nodding, etc.) (3)
3. giving subordinate the opportunity to express himself/herself; not interrupting (8)
4. noticing relevant (non) verbal hints and responding to them (1)
5. asking for approval of subordinate (3)
6. checking if his/her message correctly reaches subordinate (3)
7. building on the ideas suggested by subordinate (2)
8. communicating in a style that matches the content of message (1)
9. summarizing regularly (7)
10. paraphrasing the ideas of subordinate (3)
11. showing consideration for subordinate's situation (14)

12. taking actions that show consideration for the feelings and needs of subordinate (1)
13. asking questions regarding the subordinate's feelings (4)
14. being able to show his/her own feelings (4)
15. involving subordinate into problem solving process (5)
16. being able to discuss interpersonal conflicts without becoming upset or angry (3)
17. remaining polite to the subordinate (5)
18. treating subordinate with respect and tact (1)
19. dropping potential "no! yes!"- issues (1)
20. emphasizing also the strengths of subordinate (4)
21. neither using offensive language nor curt remarks to criticize subordinate (8)
22. being able to convey a (negative) message in a positive style (3)
23. seeing and approaching subordinate as a peer (2)
24. considering the tone of the talk and the good relationship with subordinate (2)
25. being patient towards subordinate (1)
26. coming back at what subordinate said (1)

Planning and Organizing

1. proposing a method to structure the interview (6)
2. formulating the goals and the broader framework of the interview (5)
3. conducting the interview in a structured and organized way (not jumping from one subject to another) (3)
4. setting a concrete agenda (9)
5. managing the scarce time properly (4)
6. asking subordinate not to interrupt when he/she is talking (1)
7. intervening to monitor the flow of interview (e.g., when trivialities are discussed) (5)
8. making concrete and specific agreements (17)
9. indicating what should be done in both short run and the long run (6)
10. formulating concrete deadlines (1)
11. making follow-up agreements to control and monitor subordinate's performance (18)
12. making beforehand a checklist (what should be done, how will it be measured) (1)
13. repeating the agreements made at the end (1)
14. being able to give clear-cut feedback (4)
15. proposing realistic and feasible suggestions (3)
16. noting if people, procedures, and financial resources are needed to meet objectives (4)
17. eliminating obstacles which could preclude accomplishment of agreements made (2)
18. being able to prioritize the various agreements made (5)

Note. The numbers between brackets indicate how often the behavioral incident was mentioned. Because 20 assessors generated incidents, the maximum number is 20.

APPENDIX 2

Complete Script of Candidate 1 in Role-Play with Disgruntled Subordinate.

Candidate 1	Role-player
Hello, David, take a seat. How are you?	
If you ask me, you do not sound very enthusiast. What's wrong?	Mmm, fine.
Hmm, hmm, and therefore, lately, you are a bit ...	Well, I've been very busy. These last three months, there have been a lot of new accounts: Onkion and Heys. Multivision was also a difficult one but in the end I managed it.
I understand. What do you suggest?	At this moment it's not going as smooth as in the past. I've been doing this job for almost three years. I gained a lot of new customers. Now things are going somewhat difficult.
What do you mean?	What I suggest? That the workload be reduced. It's been a hell around here, these last months.
Sorry, I had to make this clear to you, right from the beginning. I would like to discuss the course of things with you. I still find you a very good salesman but	Well, new salespeople have to be hired. More salespeople. By the way, what kind of meeting is this?
I was astonished to receive a letter from one of our key accounts, Fyrens. He is not pleased and even suggests cancelling the deal. What has happened?	So? What's the matter?
	I am also surprised to hear this. Problems with Fyrens? I thought I had a very good ..., we were so close to a new deal.

David, I understand you are upset but Fyrens is not at all pleased with how things are going right now. What have you done recently ?

Nothing at all ... ?

Hmm, hmm. Perhaps, we have found the bottleneck. If you are about to make the deal with Fyrens, than you should proceed very quickly. You know Fyrens, don't you ?

How do we resolve this issue, together ?

In my opinion, when the deal was almost done, and surely when the client was as important as Fyrens, then you should have monitored it more closely. Fyrens requires you to stick to your words. Do you follow me on that ?

Well, you have to clear the sky. In other words, resolve the problem as soon as possible. That's what really counts now. Selling is less important. You have to put our relationship with Fyrens again on the right track. What do you think ?

David, I am convinced that you will succeed. In the past, you have proven this again and again. Our partnership with Fyrens should remain very good. Euh, before I forget, David. Don't you think you should manage your time a little bit better ? This is really important for a salesman, isn't it ?

Is everything clear to you ?

As long as I remember. Euh, in fact, nothing, at least lately.

Since I last went there, I haven't talked to them.

Yes, I know them very well.

What do you suggest ?

Yes, of course, I should have monitored everything much more closely but what do I have to do exactly from now on ?

All right. I will do my best and talk to Fyrens. To be honest, I was going to talk to him this week. I am sure everything will work out just fine.

Yeah, it is.

Yes.

Fine. Well, success, David.