Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection Lee Kong Chian School Of Business

Lee Kong Chian School of Business

6-2011

The use of role-player prompts in assessment center exercises

Eveline SCHOLLAERT
Ghent University

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 Industrial and Organizational Psychology Commons, Industrial Organization Commons, and the Organizational Behavior and Theory Commons

Citation

The Use of Role-Player Prompts in Assessment Center Exercises

Eveline Schollaert and Filip Lievens

Department of Personnel Management, Work and Organizational Psychology, Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium. eveline.schollaert@ugent.be

So far, a substantial amount of assessment center (AC) studies have aimed to improve the quality of the AC method by focusing on the assessors. However, systematic studies about the role-player in AC exercises are nonexistent. This is surprising as the role-player might serve as a key figure for consistently evoking job-relevant behavior across candidates. Therefore, this study focused on the 'role' of role-players in ACs. We examined the effects of instructing role-players to use prompts among 233 candidates. Results suggest that role-players are able to use prompts and that their negative impact on candidates' reactions is negligible. In addition, some AC dimensions (problem solving and interpersonal sensitivity) were better measured when role-players used prompts. No effects on interrater reliability were found.

1. Introduction

In comparison to many other selection procedures, one of the main advantages of assessment centers (ACs) is that assessors have the opportunity to observe actual behavior in a simulated work setting. However, this advantage also brings some challenges into play. As 'open-ended' tasks (instead of multiple-choice tests, Ryan & Greguras, 1998) it is not always guaranteed that AC exercises enable to collect enough behavioral observations *per* dimension (Brannick, 2008; Howard, 2008; Lievens, 2008; Lievens, Tett, & Schleicher, 2009). As noted by Reilly, Henry, and Smither (1990), assessors sometimes have too few observations on which to base their ratings for some dimensions when not enough behaviors are evoked. They also often need to rely on one particular behavioral reaction (i.e., 'red hot' item) to score candidates on several dimensions (Brannick, Michaels, & Baker, 1989). This is troublesome in light of guidelines that state that AC exercises should 'reliably elicit a large number of dimension-related behaviors' (International Task Force on Assessment Center Guidelines, 2009, p. 247). This lack of observability of some dimensions might also have detrimental effects on the quality of the ratings and the feedback provided as the

observation of behaviors related to relevant dimensions typically serves as a basis for providing participants with detailed feedback about their strengths and weaknesses.

Therefore, our study recommends including situational stimuli for evoking behavior within AC exercises. This recommendation is applicable to both task-based and person-based means of eliciting behavior. In the context of task-based means, situational stimuli might be planted in AC exercises through exercise instructions, video-taped scenes, or virtual reality (Lievens *et al.*, 2009). This study focuses on the inclusion of situational stimuli via a person-based means of eliciting behavior, namely role-players. Role-players might serve as a practical means of structuring AC exercises by consistently evoking job-relevant behavior across candidates. As far as we know, systematic studies about the 'role' of the role-player in AC exercises are nonexistent. Therefore, role-players and role-player behavior constitute one of the least researched aspects of AC technology (Lievens & Klimoski, 2001; Zedeck, 1986).

In this study, we examine the impact of standardized prompts used by role-players on various 'hard' and 'soft' psychometric criteria. Standardized role-player prompts are defined as predetermined statements that a role-player consistently mentions in an AC across candidates to elicit behaviors related to specific job-related dimensions. In this study, three questions are of central importance. The first question examines whether role-players are able to use standardized prompts in AC

exercises. The second question explores the effects of the use of role-player prompts on two classical indices of measurement quality (i.e., interrater reliability and external construct-related validity). Finally, the third question focuses on the impact of the use of prompts on candidates' perceptions of the AC exercises.

2. Study background

Role-players are ubiquitously used in exercises for assessing and developing managerial talent. Examples are the use of role-players in AC exercises such as one-on-one interactions, oral presentations, fact findings, and sometimes even group discussions (Goldstein & Ford, 2002). Essentially, role-players have a different task as assessors. Whereas assessors should be discrete during AC exercises, role-players should 'involve' participants (Thornton & Mueller-Hanson, 2004). The most recent guidelines mention that the role-players 'should play objectively and consistently the role' (International Task Force on Assessment Center Guidelines, 2009, p. 248). To do this in consistent fashion across participants and across time, it is vital that role-players receive adequate role-player training (Thornton & Mueller-Hanson, 2004). In current AC practice, role-player training focuses mainly on providing potential role-players with information about their role. They learn to play the role objectively and consistently (International Task Force on Assessment Center Guidelines, 2009). However, the guidelines do not contain information of *how* the role-player should be objective and consistent. Moreover, empirical research in this area is also nonexistent. We suggest that role-players should also be used to evoke dimension-related behavior from candidates and that the use of prompts might serve as a practical means for role-players to do so.

Conceptually, the notion of standardized role-player prompts is based on trait activation theory. This theory states that behavior is explained by the interaction of individuals with the situation and their reactions on job-relevant cues in the situation (Tett & Guterman, 2000). By using prompts a situational stimulus for evoking job-relevant behavior is created. In role-player trainings, role-players are then taught to use multiple standardized prompts per dimension in a consistent fashion across candidates. These prompts are framed in a script as every conversation is different. On the one hand role-players need to follow this script as strictly as possible and are expected to use enough prompts per dimension. Yet, on the other hand they also need to play their role in a credible way. These opposite demands might put some pressure on role-players.

In light of these opposing demands that role-players face, a first important question is whether role-players are actually able to use prompts despite those opposing demands. To get insight into this question, cybernetic theory might be useful. This theory states that individuals have the power to adapt their behavior to the situation. The theory consists of four central elements (goal, feedback, comparator, and effector) (Bozeman & Kacmar, 1997). In this study's context, one of the 'goals' of the role-player is creating a certain image to the candidate to elicit dimension-related behavior. The 'feedback' the role-player receives, is the behavior of the candidate as the role-player needs to permanently evaluate whether the candidate gets sufficient opportunities to show relevant behavior (comparator). No discrepancies occur if the role-player can follow the script with prompts. However, if the role-player considers it is not possible to use sufficient prompts, the 'effector' will correct his/her way of acting. In other words, the role-player will adjust his/her behavior to the characteristics of the person in order to better perform the script. Thus, using cybernetic theory we predict that role-players will be able to structure the exercises by the use of prompts, despite the demanding and conflicting tasks inherent in a role-play.

Hypothesis 1: Role-players who attend a prompt-training, will use more dimension-related prompts than role-players who attend a training without prompts.

The second hypothesis deals with the effects of role-player prompt-training on the interrater reliability of the AC ratings made. We expect that the use of role-player prompts in AC exercises might affect the amount of candidate behavior to be observed in exercises. In turn, this might have beneficial effects on the interrater reliability of assessor ratings. This argument is based on two rationales. First, evoking more candidate behavior should increase the standardization of AC exercises. Second, the opportunity to observe and take notes on dimension-related behavior should increase the reliability of the ratings made in light of the principle of aggregation (Epstein, 1979), which states that the sum of a set of measurements is more stable than any single measurement from the set. Just as the reliability and content representation of a test increases with the addition of items from the same domain, assessing a given dimension in an AC exercise will improve with the addition of prompts to evoke behavior. Thus, this leads to the following hypothesis.

Hypothesis 2: If role-players attend a prompt-training, the interrater reliability will be higher than in the condition without a prompt-training.

Related to the external construct-related validity (e.g., Shore, Thornton, & Shore, 1990), we posit a next set of hypotheses about the correlations between the final dimensional AC ratings and the scores on conceptually related measures (e.g., general mental ability, personality)

in the two conditions. When role-players use prompts we expect that those prompts increase the opportunity for candidates to demonstrate dimension-relevant behavior and for assessors to observe this behavior, which might enhance the measurement of the AC dimensions. This might be evidenced by a higher overlap between AC ratings and conceptually related personality scores in the condition with prompts. In a similar vein, the use of prompts might increase the overlap between cognitive AC dimensions (e.g., problem solving) and general mental ability. Consistent with Whetzel, McDaniel, and Nguyen (2008), the cognitive loading of AC ratings then refers to the degree to which an AC rating (i.e., problem solving) is correlated with cognitive ability, whereas the personality loading of AC rating denotes the degree to which an AC rating (e.g., interpersonal sensitivity) correlates with a conceptually related measure of personality (e.g., agreeableness). Specifically, we considered problem solving to be conceptually related to general mental ability. We further expected interpersonal sensitivity, planning and organizing, and tolerance for stress to be conceptually related to agreeableness, conscientiousness, and emotional stability, respectively (Lievens, Chasteen, Day, & Christiansen, 2006). This leads to the following hypotheses.

Hypothesis 3a: If role-players attend a prompt-training, the correlation between problem-solving ratings during the AC and cognitive ability will be higher than in the condition without a prompt-training.

Hypothesis 3b: If role-players attend a prompt-training, the correlation between interpersonal sensitivity ratings during the AC and agreeableness will be higher than in the condition without a prompt-training.

Hypothesis 3c: If role-players attend a prompt-training, the correlation between planning and organizing ratings during the AC and conscientiousness will be higher than in the condition without a prompt-training.

Hypothesis 3d: If role-players attend a prompt-training, the correlation between tolerance for stress ratings during the AC and emotional stability will be higher than in the condition without a prompt-training.

Our last research question focuses on the reactions of the candidates to role-players who use prompts. On the one hand, prompts might lead to more favorable applicant perceptions because candidates might appreciate that prompts aim to elicit job-related behavior. That is, the use of prompts that evoke relevant behavior might be perceived as increasing the overlap with behavior on the job. As candidates prefer job-related selection proce-

dures (Hausknecht, Day, & Thomas, 2004), we hypothesize that the use of prompts leads to higher perceptions of job relatedness.

Hypothesis 4a: If role-players attend a prompt-training, candidates will perceive the job relatedness of the AC exercise higher than in the condition without a prompt-training.

On the other hand, the use of prompts might also reduce the realism and interpersonal warmth of AC exercises because it might detract from the natural flow of the exercise. This hypothesis is based on prior research in the interview domain that has shown that the use of structure in interviews leads to less favorable candidate perceptions as compared with unstructured interviews (Conway & Peneno, 1999). So, we expect that the use of prompts might also lead to decreases in candidate perceptions of the two-way communication and interpersonal treatment of the exercise.

Hypothesis 4b: If role-players attend a prompt-training, candidates will perceive the two-way communication in the AC exercise lower than in the condition without a prompt-training.

Hypothesis 4c: If role-players attend a prompt-training, candidates will perceive the interpersonal treatment in the AC exercise lower than in the condition without a prompt-training.

3. Method

3.1. Sample and procedure

Data were collected from 233 final-year students of a large university (57.2% females, mean age = 22.8 years, $SD = 1.8$ years). Most of the participants had majors in Law and Sciences. Participants were recruited by an invitation e-mail for a preparation session on psychological testing and assessment (i.e., simulated selection situation). At the start of the session, participants were told that they could increase their experience with taking a variety of tests. This simulated selection setting lasted approximately 1 day.

The AC exercises were targeted toward entry-level managerial jobs. We purchased a specific role-play and a presentation exercise that were targeted to applicants who pursued an entry-level managerial job. The order of the AC exercises and the role-players was counterbalanced to avoid confounding. The dimensions measured were interpersonal sensitivity, organizing and planning, problem solving, and tolerance for stress as these were identified as relevant for entry-level managerial jobs. There was anecdotal evidence that participants perceived

the simulation in a similar way as real selection settings. For instance, they all reported to be nervous and afraid to take the tests. They also wore business attire.

3.2. Role-players

3.2.1. Generation of role-player prompts

To generate role-player prompts for the dimensions a pre-study was conducted. First, to ensure a collection of prompts that were actually used in AC practice seven experienced assessors (mean age = 38.6 years, $SD = 7.87$, 57% males, mean experience in selection = 13.3 years, $SD = 8.80$) were asked to report possible prompts role-players could use to evoke dimension-relevant behavior. For the role-play, 198 unique prompts were reported and for the presentation exercise they notified 76 unique prompts, totaling 274 prompts.

Second, we refined this list by reducing prompts that were (a) inappropriate, (b) too vague, (c) too concrete, and (d) redundant. After this procedure, only 67 and 20 unique prompts were left for the role-play and for the presentation exercise, respectively. However, for both exercises there were also 17 shared prompts. As such, during the second phase a list of 104 prompts was generated.

Third, these 104 prompts were presented to two other groups of assessors: eight graduate students in Industrial and Organizational Psychology (62% males, mean age = 27.1 years, $SD = 2.17$) and 12 actual experienced assessors (42% males, mean selection experience = 5.71, $SD = 7.35$). These two groups of assessors were asked to retranslate the prompts to the dimensions. If there was an agreement of at least 70% we considered the prompt to be a good cue to evoke the respective dimension. This further reduced the number of prompts for the role-play exercise and for the presentation. The final number of prompts was 39 across both exercises. An example of a prompt was 'Mention that you feel bad about it' (interpersonal sensitivity).

3.2.2. Experimental design

Nineteen role-players (58% females; mean age = 22.9 years, $SD = 1.5$ years) were randomly assigned to one of the two conditions (role-player training without prompts and role-player training with prompts). The trainer was a consultant with a degree in Industrial and Organizational Psychology and had 15 years experience in personnel selection. Both trainings took half a day and had an identical format. The first 1.5 hr consisted of a lecture wherein participants learned the content of the two exercises and their specific roles. Next, a videotape of role-player models was presented (1.5 hr). In the role-player training without prompts, the trainer introduced the videotape (of a role-player without prompts) by explaining they had to play their role objectively and consistently, following the Guidelines and Ethical

Considerations for AC operations (see International Task Force on Assessment Center Guidelines, 2009). In the training with prompts, the trainer also introduced the video (of a role-player using prompts) by explaining to play their role objectively and consistently, but added to this a demonstration of the use of dimension-related prompts. Hence, information was provided on how to use standardized prompts to evoke dimension-related behavior. In both conditions, the last part of the training (2 hr) was composed of practical exercises, feedback, and discussion.

3.3. Measures

3.3.1. Use of prompts

Four master students of Industrial and Organizational Psychology (100% females; mean age = 21.8, $SD = 0.96$) coded the videotapes of the candidates. To this end, the coders received a half-day of training. They independently wrote down the verbatim behavior of the role-players. Next, they counted the number of times a role-player used prompts to evoke dimension-related behavior. They also counted the number of interventions that could not be considered to be evoking dimension-related behavior. Inter-rater agreement (κs above .70) was satisfactory for all dimensions. Discrepancies were resolved through discussion.

3.3.2. AC ratings

Industrial and Organizational Psychology students (66.6% females; mean age = 22.3, $SD = 0.58$) served as assessors of the AC performances. To this end, they received a half-day of training. The trainer was a certified assessor with a graduate degree in Industrial and Organizational Psychology. The training program was composed of three main parts: (a) an introduction about the basics of ACs, (b) a portrayal of the four dimensions, and (c) a workshop on the observation and rating process which included a lecture, practice, and feedback.

Given the well-known reliability problems of within-exercise dimension ratings (Brannick, 2008; Lance, 2008), final dimension ratings were made. To this end, the behavioral reporting method was used (Thornton & Byham, 1982). Upon completion of the two exercises, assessors gathered to discuss their behavioral observations and independently rated the assessees on the four dimensions on a 5-point scale, ranging from 1 = poor to 5 = excellent. Each candidate was rated by two assessors.

3.3.3. Personality scales

Three personality scales (agreeableness, conscientiousness, and emotional stability) that were conceptually related to the AC ratings were assessed with items of Goldberg's (1999) International Personality Item Pool. Each scale is composed of 10 (positively worded or negatively worded) statements related to the respon-

Table 1. Means, standard deviations, and effect sizes of the number of prompts used by the role-players broken down by exercise and dimension

	Presentation							Role-play						
	With prompts (N = 100)		Without prompts (N = 103)		Total (N = 203)			With prompts (N = 100)		Without prompts (N = 103)		Total (N = 203)		
	M	SD	M	SD	M	SD	d	M	SD	M	SD	M	SD	d
Organizing and planning	.22a	.08	.10a	.11	.16	.11	1.03	.18a	.06	.05a	.03	.12	.08	1.54
	(4.03)	(1.83)	(1.36)	(1.32)	(2.67)	(2.07)		(4.49)	(1.60)	(1.16)	(.68)	(2.80)	(2.07)	
Interpersonal sensitivity	.13b	.06	.01b	.02	.07	.08	1.55	.15a	.08	.02b	.03	.09	.08	1.54
	(2.18)	(1.24)	(.10)	(.39)	(1.12)	(1.38)		(3.87)	(1.81)	(.54)	(.85)	(2.19)	(2.18)	
Problem solving	.11b	.05	.01b	.03	.06	.07	1.57	.09b	.05	.02b	.02	.06	.05	1.36
	(2.03)	(.88)	(.18)	(.48)	(1.09)	(1.61)		(2.33)	(1.19)	(.42)	(.53)	(1.36)	(1.33)	
Tolerance for stress	.12b	.06	.02b	.05	.07	.08	1.30	.05c	.03	.00c	.00	.02	.03	1.46
	(2.18)	(1.22)	(.28)	(.63)	(1.21)	(2.36)		(1.20)	(.74)	(.04)	(.24)	(.61)	(.80)	
Total	.59	.11	.14	.12	.36	.25	1.80	.47	.11	.10	.07	.28	.20	1.85
	(10.43)	(3.04)	(1.82)	(1.60)	(6.04)	(4.94)		(11.88)	(2.93)	(2.13)	(1.60)	(6.94)	(5.41)	

Note. The means and standard deviations were calculated by the proportions of prompts used by the role-players per dimension (number of prompts per dimension per exercise/total number of interactions). Positive effect sizes mean that the proportion of the number of prompts used by the role-player was higher in the condition with prompts than in the condition without prompts. Means with different subscripts (vertical) are significantly different at .01 level. Means and standard deviations within parentheses are the means and standard deviations of the raw number of prompts. Owing to technical problems in videotaping the candidates, the total number of candidates in this table is 203 instead of 233.

dent's standing on the trait. Respondents are asked to indicate how accurately each statement describes them, using a Likert-type scale ranging from 1 = *very inaccurate* to 5 = *very accurate*. A composite score was computed by summing the items related to the given trait. Internal consistency reliabilities were .83 (agreeableness), .77 (conscientiousness), and .89 (emotional stability).

3.3.4. Cognitive ability measure

To assess cognitive ability, the Advanced Progressive Matrices (APM; Set II; Raven, Raven, & Court, 1998), a widely used measure of higher order general mental ability in both research and applied settings, was administered. The APM consists of 36 items, with a time limit of 40 min. The test booklet indicated that all test–retest reliabilities were above .90.

3.3.5. Candidate perceptions

At the end of each AC exercise, candidates received a questionnaire consisting of scales from the Selection Procedural Justice Scale (Bauer *et al.*, 2001). The first scale, job relatedness (2 items), referred to the degree to which a test appears to measure content relevant to the job (α = .81 for the role-play; α = .82 for the presentation). The second scale was two-way communication (3 items) which examined candidates' perceptions about having the opportunity to discuss their opinion. The α was .75 (role-play) and .80 (presentation). The third scale of interpersonal treatment referred to the degree to which candidates have the perception of being treated with respect. This scale consisted of 4 items and its α was .85 (role-play) and .84 (presentation).

4. Results

To examine Hypothesis 1 we conducted per exercise a $19 \times 2 \times 4$ (Role-Player $\times$ Prompt-Training $\times$ Dimension) MANOVA with repeated measures on dimension.[1] Role-players were nested within role-player training. Table 1 presents means, standard deviations, and effect sizes of the number of prompts used by the role-player, broken down by exercise and dimensions. For the presentation exercise, $F(1, 186) = 404.96$, $p < .001$, and for the role-play, $F(1, 186) = 517.70$, $p < .001$, there was a significant main effect of prompt-training. Effect sizes were large, varying from 1.03 to 1.57. Only for the role-play, a significant main effect of role-player was also found, $F(17, 186) = 1.81$, $p < .05$. To assess the practical relevance of these effects, we inspected how many prompts were used. In the role-play, the average proportion of prompts was .10 (2.13 prompts) in the condition without prompts and .47 (11.88 prompts) in the condition with prompts, so the proportion of prompts increased fivefold. In the presentation, the proportion of prompts quadrupled, from .14 (1.82 prompts) to .59 (10.43 prompts). Thus, in support of Hypothesis 1, when role-players attended a prompt-training, they were able to use significantly more prompts than when they did not follow such a training. Notwithstanding, it should also be noted that approximately half of the interactions during the exercises were not prompt related as in the role-play the average proportion of prompts was .47 and in the presentation .59. Generally, however, our results showed that role-players were able to use prompts despite the often unpredictable flow of the role-play. So, Hypothesis 1 was confirmed in both exercises.

The main effect of prompts in the role-play was also qualified by a two-way interaction effect between prompt-training and dimension, $F(3, 184) = 2.71$, $p < .05$. This meant that in the role-play, the distribution of the number of prompts per dimension was different for the two conditions. The varying distribution is indicated by subscripts (Table 1). As shown, planning was elicited most in the condition *without* prompt-training, whereas planning *and* interpersonal sensitivity were evoked mostly in the condition *with* prompt-training.

Regarding Hypothesis 2, we calculated the intraclass correlations *ICC* (2,2) between the dimensional ratings of the assessors (see Table 2). All intraclass correlations were acceptable. No differences between the prompt-training condition and the condition without prompt-training were found. So, Hypothesis 2 was not supported.

Table 2. Intraclass correlations between the dimensional ratings of the assessors for both conditions

	Without prompts ($N = 116$)	With prompts ($N = 117$)
1. Problem solving	.755	.715
2. Interpersonal sensitivity	.741	.747
3. Planning	.825	.810
4. Tolerance for stress	.663	.732

Note. The correlations are intraclass correlations *ICC* (2,2).

Table 3. Correlations between the AC dimensions and the conceptually relevant traits and abilities in both conditions

	Without prompts ($N = 116$)	With prompts ($N = 117$)
1. AC Problem solving – Cognitive ability	−.003	.149
2. AC Interpersonal sensitivity – Agreeableness	−.046	.120
3. AC Planning – Conscientiousness	.144	.154
4. AC Tolerance for stress – Emotional stability	.024	.015

Note. AC = assessment center.

Concerning Hypotheses 3a, 3b, 3c, and 3d we examined the correlations between the four dimensions and cognitive ability, and the three personality scales, respectively. All correlations were small. Table 3 shows only for the dimensions problem solving and interpersonal sensitivity a higher correlation with the relevant personality traits in the condition with prompt-training than in the condition without prompt-training. So, there was only some support for Hypotheses 3a and 3b.

To examine Hypotheses 4a, 4b, and 4c we conducted a $2 \times 2 \times 3$ (Prompt-Training × Exercise × Type of Candidate Reaction) MANOVA analysis with repeated measures on the second and third factor.[1] Table 4 presents means, standard deviations, and effect sizes of the candidate reactions. First, the results indicated a main effect of candidate reaction, $F(2, 230) = 137.04$, $p < .01$, partial $\eta^2 = .54$. Consistent with Hypothesis 4a, this main effect was qualified by a two-way interaction between prompt-training and candidate reaction, $F(2, 230) = 5.66$, $p < .01$, partial $\eta^2 = .05$. No other interaction effects were found. We further investigated whether the interaction effects were in line with the different hypotheses. For Hypothesis 4a, no significant effect of prompt-training on job relatedness was found, $F(1, 231) = 0.05$, $p = .82$, $d = .02$. Related to Hypothesis 4b, we found a positive significant effect of prompt-training on two-way communication; but it was in the opposite direction, $F(1, 231) = 5.55$, $p < .05$, $d = .32$. Finally, for Hypothesis 4c no significant effect was found, $F(1, 231) = 2.15$, $p = .14$, $d = −.20$.

5. Discussion

This study contributes to the AC literature as it provides a better understanding of one of the most unexplored components of ACs, namely role-players. The first aim of our study consisted of investigating whether role-players were able to use role-player prompts in AC exercises. Generally, our results and the effect sizes indicate that attending a prompt-training substantially increases the

Table 4. Means, standard deviations, and effect sizes of the candidates' reactions per exercise

	Presentation							Role-play						
	Condition with prompts ($N = 116$)		Condition without prompts ($N = 117$)		Total ($N = 233$)			Condition with prompts ($N = 116$)		Condition without prompts ($N = 117$)		Total ($N = 233$)		
	M	*SD*	*M*	*SD*	*M*	*SD*	*d*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*d*
Interpersonal treatment	4.00a	.71	4.16a	.58	4.09	.65	−.25	4.10a	.69	4.15a	.63	4.12	.66	−.08
Two-way communication	4.11a	.61	3.92b	.66	4.01	.64	.30	4.07a	.60	3.97a	.56	4.02	.58	.17
Job relatedness	3.38b	.73	3.34c	.66	3.36	.69	.05	3.34b	.77	3.35b	.82	3.34	.79	−.01

Note. Positive effect sizes mean that the candidates' perceptions were higher in the condition with prompts than in the condition without prompts. Means with different subscripts (vertical) are significantly different at .01 level.

number of prompts used by role-players during the assessment exercises. These results support that role-players might serve as a practical means of structuring AC exercises by consistently evoking job-relevant behavior. Accordingly, use of prompts by role-players might ensure that AC exercises adhere to the AC guideline stating that 'exercises should be carefully designed to reliably elicit a large number of dimension-related behavior' (International Task Force on Assessment Center Guidelines, 2009, p. 247). Thus, we posit that the 'role' of the role-players in AC exercises should be extended, as they might use several prompts for each dimension targeted, thereby evoking more dimension-relevant behavior.

Despite this positive evidence, using prompts is not straightforward as our results show that after following a prompt-training half of the interactions were still not prompt-related. Apparently, it is hard to continuously manage, protect, and adapt the created identity to others and to the situation (Bozeman & Kacmar, 1997). The fact that playing a role in AC exercises in a standardized way (on the basis of role-player training) is not easy was also evidenced by the main effect of role-player on the number of prompts used in the role-play. In the more structured presentation exercise, such a main effect of role-player was not found.

It is also important that role-player prompts are subtle in evoking job-relevant behavior. Mischel (1973) posited that situations should not be too strong (see also Tett & Burnett, 2003). Role-players have to be permanently cautious of formulating their prompts not too strong to avoid being too suggestive. Prompts are meant to evoke dimension-related behavior but they still have to be ambiguous in order for individual differences between candidates in reacting to them to be observed. In sum, our results suggest that role-players might be able to use prompts during AC exercises but considering the several caveats, it is not straightforward. Hence, a thorough training to use prompts should be an essential component of role-player training programs.

The second aim of our study was to examine the effects of role-player prompts on interrater reliability and external construct-related validity. As AC exercises are more standardized and more behavior is elicited when prompts are used, we expected more consistency in the ratings of the assessors in the condition with prompt-training. However, our results showed no differences between both conditions in terms of interrater reliability. Future research might consider familiarizing the assessors with the prompts. In our study, the assessors were not aware of the use of role-player prompts, so prompts probably did not function as a 'hint' for them.

Although it should be noted that all correlations between AC ratings and external measures were small, the cognitive ability measure correlated more strongly with the dimension problem solving in the condition where role-players used prompts. Furthermore, the link between the trait agreeableness and interpersonal sensitivity was higher in the prompt-training condition. So, the AC *g*-loading and the AC agreeableness loading were stronger in the condition with prompt-training than in the condition without prompt-training. For the other two pairs of dimensions and personality scales, no differences were found.

The third objective of our study was to examine the impact of the use of prompts in AC exercises on candidate perceptions of these exercises. For interpersonal treatment and job relatedness, no significant effects were found. This could be due to a ceiling effect as previous research showed that candidates already react highly favorable to AC exercises (see also the means in Table 4). For two-way communication, a significant effect was reported. Contrary to our expectations, in the condition with prompts candidates had the perception of having more opportunities to give their opinion. Candidates might have considered this as prompts providing them with more opportunity to converse with the role-player. In any case, our results suggest no risk of a negative influence of the use of prompts on candidate reactions.

This study has a number of limitations. The first limitation refers to the exercises. Prompts were developed only for role-play and presentation exercises. Future studies should examine their viability for other exercises. Second, this study was not conducted in a real selection setting and the sample consisted of final-year students without AC experience. As students are not the same as real applicants, we did our utmost best to simulate an actual selection situation. For instance, students self-selected to participate in a full-blown selection situation as they completed a variety of tests and took part in actual AC exercises. They also received feedback on their performance, and there was anecdotal evidence that participants considered the simulation in a similar way as real selection settings. However, it is clear that the generalizability of our results to field settings should be examined.

Future research should examine the effects of using prompts in AC exercises on the observability of dimensions (see Lievens *et al.*, 2009). As more dimension-related behavior is evoked, we assume that the use of prompts will increase the opportunity for candidates to demonstrate relevant behavior and for assessors to observe this behavior. Further, as more behavior related to job-relevant dimensions might be triggered, we also expect that the use of role-player prompts might increase the point-to-point correspondence with the criterion.

Acknowledgement

Note

1. We controlled for exercise order. No significant main effect was found.

References

Bauer, T. N., Truxillo, D. M., Sanchez, R. J., Craig, J., Ferrara, P., & Campion, M. A. (2001). Applicant reactions to selection: Development of the selection procedural justice scale. *Personnel Psychology*, *54*, 387–419.

Bozeman, D. P., & Kacmar, K. M. (1997). A cybernetic model of impression management processes in organizations. *Organizational Behavior and Human Decision Processes*, *69*, 9–30.

Brannick, M. T. (2008). Back to basics of test construction and scoring. *Industrial and Organizational Psychology*, *1*, 131–133.

Brannick, M. T., Michaels, C. E., & Baker, D. P. (1989). Construct validity of in-basket scores. *Journal of Applied Psychology*, *74*, 957–963.

Conway, J. M., & Peneno, G. M. (1999). Comparing structured interview question types: Construct validity and applicant reactions. *Journal of Business and Psychology*, *13*, 485–506.

Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology*, *37*, 1097–1126.

Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (pp. 7–28). Tilburg, The Netherlands: Tilburg University Press.

Goldstein, I. L., & Ford, J. K. (2002). *Training in organizations*. Belmont, CA: Thomson Learning.

Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, *57*, 639–683.

Howard, A. (2008). Making assessment centers work the way they are supposed to. *Industrial and Organizational Psychology*, *1*, 98–104.

International Taskforce on Assessment Centers Guidelines. (2009). Guidelines and ethical considerations for assessment center operations. *International Journal of Selection and Assessment*, *17*, 243–253.

Lance, C. E. (2008). Why assessment centers (ACs) don't work the way they're supposed to. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *1*, 84–97.

Lievens, F. (2008). What does exercise-based assessment really mean? *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *1*, 117–120.

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

Lievens, F., & Klimoski, R. J. (2001). Understanding the assessment center process: Where are we now? *International Review of Industrial and Organizational Psychology*, *16*, 246–286.

Lievens, F., Tett, R. P., & Schleicher, D. J. (2009). Assessment centers at the crossroads: Toward a reconceptualization of assessment center exercises. In J. J. Martocchio, & H. Liao (Eds.), *Research in personnel and human resources management* (pp. 99–152). Bingley: JAI Press.

Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review*, *80*, 252–283.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Advanced progressive matrices manual* (1998 ed.). Oxford, UK: Oxford Psychologists Press.

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., & Greguras, G. (1998). Life is not multiple choice: Reactions to the alternatives. In M. D. Hakel (Ed.), *Alternatives to traditional assessment* (pp. 183–202). Mahwah, NJ: Lawrence Erlbaum Associates.

Shore, T. H., Thornton, G. C., III, & Shore, L. M. (1990). Construct validity of two categories of assessment center ratings. *Personnel Psychology*, *43*, 101–116.

Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology*, *88*, 500–517.

Tett, R. P., & Guterman, H. A. (2000). Situation trait relevance, trait expression, and cross-situational consistency: Testing a principle of trait activation. *Journal of Research in Personality*, *34*, 397–423.

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

Thornton, G. C., III, & Mueller-Hanson, R. A. (2004). *Developing organizational simulations: A guide for practitioners and students*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Whetzel, D. L., McDaniel, M. A., & Nguyen, N. T. (2008). Subgroup differences in situational judgment test performance: A meta-analysis. *Human Performance*, *21*, 291–309.

Zedeck, S. A. (1986). A process analysis of the assessment center method. *Research in Organization*, *8*, 259–296.