

Singapore Management University Institutional Knowledge at Singapore Management University

Research Collection Lee Kong Chian School Of
Business

Lee Kong Chian School of Business

1-2011

Reshaping exercise design in assessment centers: Theory, practice, and research

Filip LIEVENS

Singapore Management University, filiplievens@smu.edu.sg

Eveline SCHOLLAERT

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 and SCHOLLAERT, Eveline. Reshaping exercise design in assessment centers: Theory, practice, and research. (2011). *Assessment centres and global talent management*. 47-60. Research Collection Lee Kong Chian School Of Business.

Available at: https://ink.library.smu.edu.sg/lkcsb_research/5807

This Book Chapter 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 libIR@smu.edu.sg.

UK.

CHAPTER

4 *Adjusting Exercise Design in Assessment Centers: Theory, Practice, and Research*

FILIP LIEVENS AND EVELINE SCHOLLAERT

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

Upon looking at this title, an obvious question is whether assessment center (AC) exercises need to be designed differently from what was done in the past. AC exercises have been around since World War I and their longevity attests to their success. So why propose some changes to this monument in personnel selection and development?

Let us first acknowledge that we concur that ACs are in still in good shape. However, every good brand needs some adjustments (which is a more appropriate term than “changes” in this context) once in a while, based on new theoretical insights and empirical research. Therefore we build our novel exercise design approach on recent insights in person-situation interactionism to make a good tool even better.

The structure of this chapter is as follows: we start with delineating the reasons behind our revised exercise design approach in ACs. Next we explain the theory behind the revised exercise design approach. In a third section, we report on our program of research related to this new approach. The fourth part discusses some other possible applications in the AC domain that are congruent with this revised exercise design approach.

Why Adjust AC Exercise Design?

Generally, ACs show a record of success: first, meta-analytic research confirmed that AC ratings were predictive of a variety of criterion measures. The meta-analysis of AC criterion-related validity studies at the dimension level (Arthur et al. 2003) contained 258 validity coefficients from 34 studies. In each study, the criterion was job-related (for example, job performance ratings, promotion, and salary). On the basis of a set of 6 dimensions Arthur et al. found a range of estimated true criterion-related validities from .25 to .39, indicating the predictive power of ACs. Second, applicants react positively to ACs. The meta-analysis of Hausknecht, Day and Thomas (2004) showed that behavior sample-based selection procedures were perceived more favorably than, for instance, cognitive ability tests and personality inventories. In particular applicants view ACs as more face valid than most

1 other selection procedures. Third, a recent meta-analysis demonstrated that different 1
 2 assessors tend to agree when they are asked to evaluate candidates (Connelly and Ones 2
 3 2008). Fourth, the majority of AC studies on adverse impact attest to the widely held 3
 4 view that ACs are reasonably unbiased regarding race and gender. According to the meta- 4
 5 analysis of Dean, Roth and Bobko (2008), positive results were reported, in particular 5
 6 for females and Hispanics. For Blacks, however, their results suggested that ACs may be 6
 7 associated with more adverse impact than was previously thought in the literature, but 7
 8 still have less adverse impact than the typical cognitive ability test. 8

9 Besides these benefits leading to the popularity of ACs, one of the main advantages 9
 10 of ACs is that assessors have the opportunity to observe actual behavior in a simulated 10
 11 work setting. This key focus on behavior in ACs is also well reflected in the most recent 11
 12 *Guidelines and Ethical Considerations of Assessment Center Operations*, in which the 12
 13 observation of overt behavioral responses is described as a necessary and fundamental 13
 14 component of ACs (International Task Force on Assessment Center Guidelines 2009). 14
 15 The *Guidelines* further state that AC designers should attempt to design exercises that 15
 16 evoke a large number of job-related behaviors because this should give assessors enough 16
 17 opportunities to observe job-related behavior. Generating more job-related behaviors 17
 18 in an exercise is quintessential for developmental ACs, because these behaviors serve 18
 19 as a basis for providing participants with detailed developmental feedback about their 19
 20 strengths and weaknesses. The more behavioral examples one can assemble, the more 20
 21 convincing the feedback will likely be. In addition, regardless of how that behavior is 21
 22 then captured and evaluated by assessors (for example, in exercise/task-based models, 22
 23 dimension-based models), eliciting and observing behavior is key to effective assessment 23
 24 and development centers. 24

25 Although the AC *Guidelines* and the literature emphasize the importance of exercises 25
 26 providing sufficient opportunities for observing job-related behavior, various authors have 26
 27 emphasized that AC exercises score not that well in terms of observability of behavior: 27

- 28 29 • First, Bycio, Alvares, and Hahn (1987: 472) noted that “assessors within an exercise 29
 30 are sometimes, if not usually, forced to base *all* of their judgments on *four or five* 30
 31 behaviors.” 31
- 32 • Next, Brannick, Michaels, and Baker (1989) mentioned that assessors often need to 32
 33 rely on *one* particular behavioral reaction (“red hot” item) to score candidates on 33
 34 *several* dimensions. 34
- 35 • Third, Reilly, Henry and Smither (1990) pointed out that assessors sometimes have 35
 36 too few observations on which to base their ratings for some dimensions, when not 36
 37 enough behaviors are evoked. 37
- 38 • Furthermore, Kudisch, Ladd and Dobbins (1997) suggested that consistency of AC 38
 39 ratings across exercises may be enhanced when dimensions are easier to observe. For 39
 40 example, the dimension Communication (which is overt in most exercises) produced 40
 41 more convergent validity, as opposed to Problem Analysis (which is less observable 41
 42 in most exercises). 42
- 43 • Although the meta-analysis of Connelly and Ones (2008) revealed good inter-rater 43
 44 reliability coefficients for AC ratings, reliability was lowest for so-called within- 44
 45 exercise dimension ratings (ratings made on one dimension within a specific exercise). 45
 46 The main argument is that such ratings are often based on rather limited behavioral 46
 47 evidence. 47

1 In sum, it seems that it is not always guaranteed that AC exercises enable assessors to 1
 2 collect enough behavioral observations per dimension (see Brannick 2008, Howard 2
 3 2008, Lievens 2008, Lievens, Tett and Schleicher 2009). An explanation can be found 3
 4 in the traditional AC paradigm which focuses on the exercise as a whole. In this holistic 4
 5 approach, the exercise as a whole is seen as a vehicle for evoking behavior (Howard 5
 6 2008, Lievens et al. 2009, McFarland et al. 2005). However, as noted above, research 6
 7 has shown that this holistic exercise approach might occasionally be problematic in 7
 8 that an insufficient number of behaviors are elicited. That is the reason why we argue 8
 9 for a more molecular approach by planting situational stimuli within exercises to 9
 10 enhance the observability of dimensions across a variety of dimensions in ACs. Upfront 10
 11 we acknowledge that practitioners might already have intuitively attempted to elicit 11
 12 dimension relevant behavior by building context, content, personnel, problems into 12
 13 exercises or by instructing role players to use prompts in response to paths chosen by 13
 14 participants. However, our point is that theory and research has not shown systematically 14
 15 how the behavioral elicitation process works and has not provided empirical evidence of 15
 16 its effectiveness 16

17
 18

19 Theoretical Background 19

20

21 It is generally acknowledged that behavior of candidates in ACs is determined neither 21
 22 solely by dispositional factors (stable personal characteristics of candidates) nor solely by 22
 23 situational factors (AC exercises) but by the interaction of the person and the situation. 23
 24 Therefore, it is relevant to conceptualize the occurrence of candidate behavior in ACs in 24
 25 terms of a recent interactionist theory such as trait activation theory (Lievens et al. 2009; 25
 26 Tett and Burnett 2003). Trait activation theory focuses on the person-situation interaction 26
 27 to explain behavior based on responses to trait-relevant cues found in situations (Tett and 27
 28 Guterman 2000). These observable responses serve as the basis for behavioral ratings in a 28
 29 variety of assessments such as ACs (Tett and Burnett 2003). 29

30 According to trait activation theory, two factors are important to understand in 30
 31 which situations a trait is likely to manifest itself in behavior. First, trait activation 31
 32 theory emphasizes the importance of *situation trait relevance*. A situation is considered 32
 33 relevant to a trait if it provides cues for the expression of trait relevant behavior (Tett 33
 34 and Guterman 2000). Thus, situation trait relevance is a qualitative feature of situations 34
 35 that is essentially trait specific; it is informative with regard to which cues are present 35
 36 to elicit behavior for a given latent trait. Relatedly, trait activation theory states that 36
 37 situations should provide ample opportunities for behavior to be expressed. This idea 37
 38 builds on the well-known principle of aggregation (Epstein 1979) in social psychology, 38
 39 which states that the sum of a set of measurements is more stable than any single 39
 40 measurement from the set. For example, when someone is having an accident and is 40
 41 confronted with an angry driver, this situation provides cues for traits such as Emotional 41
 42 Stability. Conversely, this situation is less relevant to evoke traits such as Imagination 42
 43 (Openness). 43

44 *Situation strength* is the second relevant factor from the trait activation perspective. 44
 45 Situation strength is more of a continuum that refers to how much clarity there is with 45
 46 regard to how the situation is perceived. Strong situations involve unambiguous behavioral 46
 47 demands and are therefore likely to negate almost all individual differences in behavior 47

1 without regard to any specific trait. Conversely, weak situations are characterized by more 1
 2 ambiguous expectations, enabling much more variability in behavioral responses to be 2
 3 observed (Meyer, Dalal and Hermida 2010). For instance, at the end of a busy day a 3
 4 shop assistant may be confronted with a messy shop full of odds and ends left by the 4
 5 customers. When the supervisor instructs the shop assistant to clean the mess in the 5
 6 shop, it will be much more difficult to observe individual differences related to the trait 6
 7 order, whereas the opposite might be true in the absence of such clear-cut supervisory 7
 8 instructions (without instructions some shop assistants will immediately start to clean 8
 9 the shop, others will not notice the mess). 9

10 Thus, trait activation theory has key implications for AC exercises (see Lievens et 10
 11 al. 2009). That is, the application of trait activation theory involves recognition of the 11
 12 importance of building multiple stimuli into the AC exercises. Accordingly, exercises can 12
 13 be explicitly designed to increase their situation trait relevance. In this respect, Brannick 13
 14 (2008: 132) cogently argued to “deliberately introduce multiple dimension-relevant 14
 15 items or problems within the exercise and to score such items.” Apart from increasing 15
 16 the situation trait relevance of AC exercises, trait activation theory also suggests taking 16
 17 situation strength into account when planting stimuli in AC exercises. That is, behavior 17
 18 elicitation should avoid presenting the candidate with a too strong situation (in terms of 18
 19 behavioral demands). For example, role players might create a relatively weak situation by 19
 20 showing for a moment a sad facial expression (prompt to evoke Interpersonal Sensitivity). 20
 21 Some candidates will ask what is bothering the role player, whereas other candidates will 21
 22 ignore the expression or even will not notice the expression, leading to variability in 22
 23 candidate reactions. On the other hand, to evoke Interpersonal Sensitivity the role player 23
 24 might also start to sob. Almost every candidate will notice this and will react on it. This 24
 25 is probably too strong of a situation so that variability in candidate reactions will be 25
 26 masked. 26

27

28

29 Examples of Situational Stimuli 29

30

31 In the previous section, we outlined some general principles for eliciting candidate 31
 32 behavior in ACs. In this section, we provide five different examples of how to put this 32
 33 general logic into practice. 33

34 The first approach entails adapting the content of the exercise. Let us take an oral 34
 35 presentation with challenging questions as an example. Examples of stimuli to elicit 35
 36 behavior relevant to a dimension such as Resistance to Stress (a facet of the broader trait 36
 37 of Emotional Stability) might be the inclusion of a stringent time limit, sudden obstacles, 37
 38 or information overload. In a more systematic way, AC designers might ensure that 38
 39 several content cues are embedded at the task, social, and organizational levels within a 39
 40 given exercise (if job-related, of course). 40

41 A second way to elicit job-related behavior is through exercise instructions. In ACs, 41
 42 exercise instructions provide information and set expectations for candidates about what 42
 43 behavior to show or not to show. For example, exercise instructions might be vague (for 43
 44 example, “solve the problem”) or more concrete (for example, “motivate the problem 44
 45 subordinate”). Similarly, exercise instructions might be unidimensional (for example, 45
 46 reach consensus) or multidimensional (for example, reach consensus and make the 46
 47 company more profitable). To date, we know little about how such exercise instruction 47

1 variations might affect the behavior demonstrated, in terms of either direct effects or in 1
2 interactions with underlying personality traits. 2

3 Thirdly, when interpersonal exercises are used, role-player cues are an additional 3
4 means of eliciting job-related behavior. In current AC practice, role players are typically 4
5 given a specific list of things to do and to avoid. Role players are also trained to perform 5
6 realistically and consistently across candidates. Although these best practices have 6
7 proven their usefulness over the years, a key function of trained role players consists of 7
8 evoking behavior from candidates (Thornton and Mueller-Hanson 2004). Trait activation 8
9 theory can help identify which specific behaviors might be evoked by specific role-player 9
10 stimuli (prompts). Prompts are defined as predetermined statements that a role player 10
11 consistently mentions in an AC across candidates to elicit behaviors related to specific job- 11
12 related dimensions. For example, to arouse behavior related to Interpersonal Sensitivity, 12
13 the role player might state that he feels bad about a candidate's decision. Similarly, 13
14 role players might trigger behavior related to Planning and Organizing (deeper trait of 14
15 Conscientiousness) by asking how the candidate will implement his or her solution. 15

16 It is important that these role-player cues should *subtly* elicit assessee behavior 16
17 because the situations might otherwise become too strong. Indeed, role-player prompts 17
18 might vary from being very explicit (strong) to being very implicit (weak) in eliciting the 18
19 dimensions targeted. We illustrate this notion of situation strength in role-player prompts 19
20 again with the dimension of Interpersonal Sensitivity. That is, role players might react 20
21 to a decision made by the candidate by showing momentarily a distressed expression on 21
22 their face (weak situation) or might start to sob (extremely strong situation). 22

23 Fourthly, one might consider including a large number of shorter exercises (exercise 23
24 "vignettes") in the AC. For example, Brannick (2008) recommends using five 6-minute role 24
25 plays instead of a single 30-minute role play (for example, with a problem subordinate) so 25
26 that one obtains samples of performance on a large number of independent tasks that are 26
27 each exclusively designed to elicit behavior related to a specific trait (see also Motowidlo, 27
28 Hooper and Jackson 2006, for the use of 1- or 2- minute role plays). As another example, 28
29 one could aim to measure communication by including "speed" role plays with a boss, 29
30 peers, colleagues, customers, and subordinates. 30

31 Finally, stimuli could also be presented via videotape, PC, or even virtual reality. In 31
32 the videotape approach, resembling earlier social intelligence measures (Stricker and 32
33 Rock 1990), candidates are shown short scenes and asked to react to what they saw. 33
34 Recent applications even enable creation of avatar-based simulation exercises wherein 34
35 participants take on a virtual identity and are confronted with standardized stimuli in a 35
36 virtual workplace (Rupp, Gibbons and Snyder 2008). 36

37 37

38 38

39 **Overview of Empirical Research** 39

40 40

41 Let us start by acknowledging that some earlier studies have already scrutinized the effects 41
42 of specific characteristics of AC exercises (Highhouse and Harris 1993, Schneider and 42
43 Schmitt 1992). For instance, Schneider and Schmitt (1992) experimentally manipulated 43
44 the effects of exercise content (competitive vs. cooperative demands) and exercise form 44
45 (for example, role play vs. group discussion) on candidate ratings. The form of the 45
46 exercise emerged as the most important exercise factor in leading candidates to perform 46
47 differently across exercises. 47

1	So, the focus of this limited number of prior studies on AC exercise characteristics	1
2	has typically been the <i>exercise as a whole</i> . Specific situational stimuli <i>within exercises</i> were	2
3	not investigated. To fill this gap in empirical research and to put our aforementioned	3
4	theory to the test we set up a research program. So far, the following questions have been	4
5	addressed:	5
6		6
7	1. Is it possible to build situational stimuli in AC exercises?	7
8	2. What are the effects of building situational stimuli in AC exercises on observability?	8
9	3. What are the effects of building situational stimuli in AC exercises on inter-rater	9
10	reliability?	10
11	4. What are the effects of building situational stimuli in AC exercises on construct-	11
12	related validity?	12
13	5. What are the effects of building situational stimuli in AC exercises on applicant	13
14	reactions?	14
15		15
16	The remainder of this section provides an overview of the available empirical research	16
17	evidence. Note that most of the studies are still ongoing and that only preliminary	17
18	research evidence is presented.	18
19		19
20	IS IT POSSIBLE TO BUILD SITUATIONAL STIMULI IN AC EXERCISES?	20
21		21
22	Some situational stimuli (for example, exercise instructions, video-based stimuli or stimuli	22
23	in virtual reality) can be used and implemented independently of the candidate. In that	23
24	case, it is relatively straightforward that the answer to this question is "Yes." However,	24
25	other situational stimuli (role-player prompts) are given in a constantly changing	25
26	situation. In that case, the answer to this question might be more complex. Prompts	26
27	create a situational stimulus for evoking job-relevant behavior. Role players are taught	27
28	to use multiple standardized prompts for each dimension in a consistent fashion across	28
29	candidates. These prompts provide a framework for responding as every conversation is	29
30	different. By using prompts a situational stimulus for evoking job-relevant behavior is	30
31	created. In role-player training, role players are then taught to use multiple standardized	31
32	prompts per dimension in a consistent fashion across candidates. These prompts are	32
33	framed in a script as every conversation is different. On the one hand role players need	33
34	to follow this script as strictly as possible, thereby being expected to use enough prompts	34
35	per dimension. Yet, on the other hand they also need to play their role in a credible	35
36	way. Each candidate reacts differently so that the role player often has to pursue several	36
37	dissimilar strategies to maintain the script. These opposing demands might put some	37
38	pressure on role players. Consequently, the question is whether role players are actually	38
39	able to use prompts in a standardized way despite those opposing demands.	39
40	Schollaert and Lievens (in press) sought to examine this. Their focal question was	40
41	whether role players were able to use prompts. A sample of role players was randomly	41
42	assigned to one of the following two conditions: role-player training without prompts	42
43	and role-player training with prompts. Generally, the results indicated that attending	43
44	training with prompts substantially increased the number of prompts used by role	44
45	players during the assessment exercises. Effect sizes were large. In the role-play exercise,	45
46	the average proportion of prompts increased fivefold. In the presentation exercise, the	46
47	proportion of prompts quadrupled. Thus, these results support the view that role players	47

1 might serve as a practical means of structuring AC exercises to consistently evoke job- 1
 2 relevant behavior. Despite this positive evidence, using prompts is not straightforward 2
 3 as results also showed that after attending role-player training with prompts, half of the 3
 4 interactions still failed to show any prompts. 4

5 6 WHAT ARE THE EFFECTS OF BUILDING SITUATIONAL STIMULI INTO 6 7 AC EXERCISES ON OBSERVABILITY? 7

8
 9 As noted above, the *AC Guidelines* have underscored the importance of the observability 9
 10 of behavior. Hence, an important argument for using situational stimuli is based on the 10
 11 key assumption that they should increase the number of observations per dimension 11
 12 noted down by assessors. 12

13 Schollaert and Lievens (2009) tested this assumption of the increased situational 13
 14 relevance of AC exercises by contrasting two vehicles for increasing behavior observability, 14
 15 namely instructions to role players prior to a role play and role-player prompts during 15
 16 the role play. No main effect was found for exercise instructions. Apparently, providing 16
 17 specific exercise instructions did not influence the number of good observations. However, 17
 18 results showed a main effect for prompt-training, with the use of role-player prompts 18
 19 leading to greater observability of dimension-relevant behavior. Thus, the inclusion of 19
 20 situational stimuli and especially the use of prompts might serve as a practical vehicle 20
 21 to avoid that assessors need to rely on too few behavioral reactions to score candidates 21
 22 (Brannick et al. 1989). 22

23 24 WHAT ARE THE EFFECTS OF BUILDING SITUATIONAL STIMULI IN AC EXERCISES 24 25 ON INTER-RATER RELIABILITY? 25

26
 27 If situational stimuli are built into AC exercises, one might expect that this has beneficial 27
 28 effects on inter-rater reliability for several reasons. First, evoking more candidate behavior 28
 29 should increase the standardization of those exercises. Second, the opportunity to observe 29
 30 and take notes on dimension-related behavior should also increase the reliability of the 30
 31 ratings made in light of the aforementioned principle of aggregation (Epstein 1979). Just 31
 32 as the reliability and content representation of a test increases with the addition of items 32
 33 from the same domain, assessing a given dimension in an AC exercise might improve 33
 34 with the addition of dimension-specific cues. Third, the use of standardized situational 34
 35 cues in ACs can be compared to the use of standardized questions among interviewers. 35
 36 Research in the interview domain has shown that the inter-rater reliability of structured 36
 37 interviews is higher than that of unstructured interviews (Conway, Jako and Goodman 37
 38 1995). Thus the use of standardized dimension-related stimuli across candidates might 38
 39 increase the standardization, the structure, and the consistency of AC ratings. 39

40 In our research, we found empirical support for these hypotheses only when assessors 40
 41 were also familiar with the situational stimuli built into the AC exercises. For example, 41
 42 Lievens, Keen, and Schollaert (2010) compared three conditions. In the low behavior 42
 43 elicitation condition, no formal attempts were implemented to evoke dimension-related 43
 44 behavior. In the medium behavior elicitation condition, role players were trained to use 44
 45 specific prompts for evoking candidate behavior. The high behavior elicitation condition 45
 46 was similarly designed as the medium behavior elicitation condition, with the addition 46
 47 that assessors were also familiarized with the prompts for eliciting behavior. In that 47

1 condition, assessors knew which prompts were related to which dimensions. Inter-rater 1
 2 reliability was highest in the third condition, where role players used prompts for evoking 2
 3 behavior and where assessors were also familiar with the prompts used by role players. 3

4 Taken together, these results make sense as they indicate that it is not enough that 4
 5 exercises generate more behavior. In addition, it seems important that assessors receive 5
 6 information about the cues that elicit behavior in order to consistently observe, classify, 6
 7 and rate that behavior. 7

8 9 WHAT ARE THE EFFECTS OF BUILDING SITUATIONAL STIMULI IN AC EXERCISES 9 10 ON CONSTRUCT-RELATED VALIDITY? 10

11
 12 When situational stimuli are used to elicit a higher number of behaviors in AC exercises, 12
 13 it can be expected that dimensions are also better measured in AC exercises. Thus, one 13
 14 might anticipate beneficial effects on the construct-related validity of AC exercises. Given 14
 15 the higher number of behaviors available per dimension, assessors should be less prone 15
 16 to using a couple of behavioral items per dimension to rate these dimensions. In other 16
 17 words, their ratings should be less susceptible to halo. The convergence of their ratings 17
 18 across exercises might also increase due to the higher observability of the dimensions. 18
 19 This might be an important advantage of the use of situational stimuli as the construct- 19
 20 related validity of AC ratings has traditionally been identified as one of the weaker points 20
 21 of AC technology, especially when within-exercise dimension ratings are used (Bowler 21
 22 and Woehr 2006; Lance 2008; Lievens and Conway 2001). 22

23 In recent years, some studies have put these expectations to the test. Schollaert and 23
 24 Lievens (in press) found that Problem-solving and Interpersonal Sensitivity dimensions 24
 25 were better measured in AC exercises when role players used prompts for evoking these 25
 26 dimensions. This was evidenced by higher correlations between Problem-solving and 26
 27 Interpersonal Sensitivity ratings and Cognitive Ability and Agreeableness, respectively. 27
 28 Lievens et al. (2010) focused on behavior elicitation via role-player prompts in a sample 28
 29 of actual candidates for a managerial job. As noted above, they distinguished between 29
 30 three levels of behavior elicitation (high, medium and low). Results showed that 30
 31 construct-related validity (convergent and discriminant correlations) was highest in the 31
 32 high behavior elicitation condition. That is, significantly more evidence for convergent 32
 33 and discriminant validity was established when role players used prompts for eliciting 33
 34 behavior and when assessors were familiar with these prompts. 34

35 Two other studies experimented with the use of video-based vignettes for eliciting 35
 36 dimension-related behavior. In Lievens (2009), candidates for police-officer jobs watched 36
 37 video-based scenes. Each of these scenes triggered a specific dimension. At the end of 37
 38 each scene, the character in the video spoke directly into the camera. Candidates were 38
 39 next required to answer the character directly, with their verbal and non-verbal reply 39
 40 being captured by a webcam. These reactions were then coded by trained assessors. One 40
 41 set of analyses examined the consistency of assessors' dimensional ratings across scenes 41
 42 (convergent validity). That is, did scenes that triggered a similar dimension provide a 42
 43 consistent measurement of that specific dimension? In line with expectations, the 43
 44 consistency in assessor ratings was acceptable (only a more ambiguous dimension such 44
 45 as Integrity scored a bit lower), confirming that the use of multiple videotaped scenes for 45
 46 measuring one dimension might serve as a good vehicle for obtaining a more consistent 46
 47 measurement of the targeted dimension. Brink et al. (2008) also showed candidates short 47

1 video scenes and asked them to react to what they saw. They focused on the discriminant 1
 2 validity of assessor ratings and found that assessors were able to make good differentiations 2
 3 among the various dimensions. 3

4 5 WHAT ARE THE EFFECTS OF BUILDING SITUATIONAL STIMULI IN AC EXERCISES 5 6 ON APPLICANT REACTIONS? 6

7
 8 Traditionally, AC exercises are held in high regard by participants. However, the use of 8
 9 situational stimuli might also have an impact on how participants perceive AC exercises. 9
 10 On the one hand, the use of situational stimuli might lead to more favorable applicant 10
 11 perceptions because candidates might appreciate that prompts aim to elicit job-related 11
 12 behavior. That is, the use of stimuli that evoke relevant behavior might be perceived as 12
 13 increasing the overlap with behavior on the job. As candidates prefer job-related selection 13
 14 procedures (Hausknecht et al. 2004) the inclusion of situational stimuli might lead to 14
 15 higher perceptions of job-relatedness. 15

16 On the other hand, the use of situational stimuli might also have some negative 16
 17 effects on applicant perceptions. Possibly it reduces the realism and interpersonal warmth 17
 18 of AC exercises because it might detract from the natural flow of the exercise. In fact, prior 18
 19 research in the interview domain has shown that the use of structure in interviews led 19
 20 to less favorable candidate perceptions as compared to unstructured interviews (Conway 20
 21 and Peneno 1999). 21

22 So far, only one study has examined the impact of the use of situational stimuli 22
 23 (in this case role-player prompts) on applicant perceptions. Schollaert and Lievens (in 23
 24 press) examined the effects of situational stimuli in the form of role-player prompts on 24
 25 perceptions of job-relatedness, two-way communication, and interpersonal treatment. 25
 26 They hypothesized that the use of prompts led to higher perceptions of job-relatedness and 26
 27 to decreases in the perceptions of two-way communication and interpersonal treatment. 27
 28 Half of the candidates were confronted with role players not using prompts and half of the 28
 29 candidates were confronted with role players using prompts. For interpersonal treatment 29
 30 and job-relatedness, no significant effects were found. This could be due to a ceiling effect 30
 31 as previous research showed that candidates already react highly favorable to AC exercises. 31
 32 For two-way communication, a significant effect was reported. Contrary to expectations, 32
 33 candidates had the perception of having more opportunities to give their opinion in the 33
 34 condition with prompts. Candidates might have considered prompts as providing them 34
 35 with more opportunity to converse with the role player. In any case, these results suggest 35
 36 that the use of prompts does not have a negative influence on candidate reactions. However, 36
 37 a caveat is warranted as this study was not conducted in a real selection setting and the 37
 38 sample consisted of final year university students without AC experience. So, future research 38
 39 with experienced applicants is needed. In addition, the effects of the use of situational cues 39
 40 on participants' acceptance of AC feedback should be scrutinized. 40

41 42 SUMMARY AND FUTURE RESEARCH 42

43
 44 So far, researchers have experimented with three types of situational stimuli: exercise 44
 45 instructions, role-player prompts, and videotaped stimuli. Generally, this overview of 45
 46 the growing empirical evidence shows that it is possible to build situational stimuli into 46
 47 AC exercises. Incorporating multiple situational stimuli is also found to generate the 47

1 anticipated effects in terms of increasing the number of behaviors to be observed. That 1
 2 is, assessors noted down more behaviors on their observation forms in the case where 2
 3 situational stimuli were built into the exercises. However, situational stimuli do not 3
 4 guarantee that assessors are taking them into account in their ratings. Only if assessors 4
 5 are familiarized with the situational cues they have effects on inter-rater reliability, 5
 6 convergent validity, and discriminant validity. Finally, applicants do not seem to notice 6
 7 any negative effects of the use of situational cues in AC exercises. 7

8 Future research should extend these results. A key issue consists of investigating 8
 9 whether the use of situational cues also increases the criterion-related validity of the AC 9
 10 exercise; the rationale being that as more dimension-relevant behavior will be elicited, 10
 11 the use of situational stimuli might increase the overlap with the criterion. 11

12 Another issue is whether the provision of cues changes the dimension that is actually 12
 13 being measured. That is, does it change from a measure of maximum performance to 13
 14 typical performance, or vice versa? The provision of cues in AC exercises may make the 14
 15 situation stronger, and suggest to participants what they should do, rather than allow 15
 16 them to choose what to do. Interestingly, the effects might differ depending on the type 16
 17 of dimensions (personality like dimensions versus ability like dimensions). For example, 17
 18 consider a role play where the candidate is a manager and the role player is a supervisor 18
 19 having a problem with an employee. With no cues, the candidate may or may not engage 19
 20 in coaching behaviors; with cues from the role player (for example, "Well, what can you 20
 21 do to help me with my problem?"), the candidate may start giving suggestions. The 21
 22 exercise then provides behavior relevant to coaching (maximum performance or "can 22
 23 do"), but at the cost of denying the candidate the opportunity of proactively displaying 23
 24 any inclination to provide coaching to the subordinate (typical performance or "will 24
 25 do"). So, giving cues might change the dimensions being measured from a measure of 25
 26 tendency to coach to a measure of ability to coach (see also McDaniel, Hartman, Whetzel 26
 27 and Grubb, 2007). 27

28 The same sort of analysis should be applied to dimensions like problem-solving. 28
 29 With minimal cues, the candidate may do very little systematic problem analysis and 29
 30 decision analysis. But the role player might give a series of cues in follow-up questions 30
 31 after the presentation: "What led you to say you would do ... ?" "What other solutions 31
 32 did you consider ... ?" The situation then may become stronger, and the candidate would 32
 33 give answers that he did not really consider in his initial preparation. Thus, one could 33
 34 evaluate the ability to do problem-solving because more of those behaviors would be 34
 35 displayed. In summary, to measure "personality-like" traits, providing cues might make 35
 36 the situation stronger and reduce individual differences in the behavior one wants to 36
 37 observe. To measure "cognitive ability - like" traits, providing cues may ensure that 37
 38 relevant behaviors are displayed, and thus enhance measurement accuracy. In future 38
 39 studies, we plan to test these ideas. 39

42 Other Implications 42

43 43
 44 So far, we have focused on the use of situational cues in AC exercises. However, the 44
 45 inclusion of situational cues not only has implications for AC exercises, but also for at 45
 46 least the following three other components of AC technology, namely the design of 46
 47 behavioral checklists, assessor training, and alternate AC exercises. 47

1	BEHAVIORAL CHECKLISTS	1
2		2
3	As noted above, initial research with the inclusion of predetermined situational cues in	3
4	AC exercises suggests that it is important that assessors are familiar with the cues used.	4
5	To accomplish this, a practical approach might consist of including the situational	5
6	cues that were designed to elicit candidate behavior (for example, the role-player	6
7	statements) in the behavioral checklists provided to the assessors. In an even more	7
8	structured format, these cues could be presented in their anticipated chronological	8
9	order, with the candidate behaviors to be observed arranged around them. Accordingly,	9
10	assessors are reminded and prompted by the situational cues when attending to	10
11	candidate behavior. It might help them to “see the forest for the trees” in the complex	11
12	stimuli triggered by AC exercises. Brannick (2008) refers to this approach as aligning	12
13	the stimulus content of the exercises with the scoring rubric (as is sometimes done	13
14	in in-baskets, wherein they are provided behaviorally anchored rating scales which	14
15	show, for each in-basket item, the types of responses that would be considered high/	15
16	medium/low performance).	16
17		17
18	ASSESSOR TRAINING	18
19		19
20	A related implication consists of familiarizing assessors with the situational cues in	20
21	assessor training. In current assessor training practice, the focus is placed on imposing	21
22	a consistent frame-of-reference on assessors (Lievens 2001). In such training programs,	22
23	the dimensions and the accompanying behaviors play logically a crucial role. However,	23
24	it is equally important that assessors know when specific behavior is potentially being	24
25	activated by various situational stimuli. We are not aware of studies that have examined	25
26	such a more comprehensive assessor training approach. Apart from teaching assessors	26
27	the cues in a lecture, other options are possible. For example, when the same individuals	27
28	serve as assessors and role players in the AC, they also learn to use the cues. This might	28
29	be especially helpful in cross cultural settings where candidates and assessors come from	29
30	different backgrounds.	30
31		31
32	ALTERNATE AC EXERCISES	32
33		33
34	One potential benefit of incorporating situational cues in AC exercises is that the AC	34
35	exercise is no longer a “black box”; that is, the development and inclusion of situational	35
36	cues within an AC exercise breaks it down into different parts and components. Thus, the	36
37	identification of situational cues might guide the determination of the deeper structural	37
38	aspects (the so-called radicals, to use a term from item generation theory; Irvine, Dann	38
39	and Anderson 1990; Lievens and Sackett 2007) of an AC exercise, in terms of providing a	39
40	template of what aspects of the exercise map onto which dimensions, and should be kept	40
41	constant across exercises.	41
42	Such a more molecular approach to AC exercise design might make it easier to	42
43	develop alternate forms of AC exercises (Brummel, Rupp and Spain 2009). For example,	43
44	one might develop several role-player cues to evoke behavior related to the dimension	44
45	of Interpersonal Sensitivity in a series of role plays. Superficial differences among the	45
46	cues would be incidental to their deeper similarities (as radicals) in targeting the same	46
47	dimension. The same might be done for other dimensions. Thus, more generally, to	47

1 construct alternate forms of AC exercises, we suggest changing the surface features of an 1
 2 AC exercise, while keeping the deep structure of the exercise intact. 2
 3 3
 4 4

5 Epilogue 5

6 6
 7 In this paper, we proposed to build multiple situational stimuli in AC exercises. We also 7
 8 reported on various studies that have implemented this approach. We want to emphasize 8
 9 that we do not suggest that current best practices of exercise development (and AC design 9
 10 in general) should be abandoned. Rather, we argue that our approach should also play 10
 11 a more prominent role in such development, with the goal of making a good tool even 11
 12 better. Whereas current practices typically simulate key task, social, and organizational 12
 13 demands of the job, we see untapped potential in planting multiple stimuli within 13
 14 exercises as a systematic and structured means of increasing the frequency and variability 14
 15 of job-related behavior in AC exercises. 15
 16 16
 17 17

18 References 18

- 19 19
 20 Arthur, W. Jr., Day, E.A., McNelly, T.L. and Edens, P.S. 2003. A meta-analysis of the criterion-related 20
 21 validity of assessment center dimensions. *Personnel Psychology*, 56, 125–154. 21
 22 Bowler, M.C. and Woehr, D.J. 2006. A meta-analytic evaluation of the impact of dimension and 22
 23 exercise factors on assessment center ratings. *Journal of Applied Psychology*, 91, 1114–1124. 23
 24 Brannick, M.T. 2008. Back to basics of test construction and scoring. *Industrial and Organizational* 24
 25 *Psychology*, 1, 131–133. 25
 26 Brannick, M.T., Michaels, C.E. and Baker, D.P. 1989. Construct validity of in-basket scores. *Journal of* 26
 27 *Applied Psychology*, 74, 957–963. 27
 28 Brink, K.E., Lance, C.E., Bellenger, B.L., Morrison, M.A., Scharlau, E.A. and Crenshaw, J.L. 2008, 28
 29 April. Discriminant validity of a “next generation” assessment center. In: B.J. Hoffman (Chair), 29
 30 *Reexamining assessment centers: Alternate approaches*. Symposium conducted at the Annual 30
 31 Conference of the Society for Industrial and Organizational Psychology, San Francisco, CA. 31
 32 Brummel, B.J., Rupp, D.E. and Spain, S.M. 2009. Constructing parallel simulation exercises for 32
 33 assessment centers and other forms of behavioral assessment. *Personnel Psychology*, 62, 137–170. 33
 34 Bycio, P., Alvares, K.M. and Hahn, J. 1987. Situational specificity in assessment center ratings: A 34
 35 confirmatory factor analysis. *Journal of Applied Psychology*, 72, 463–474. 35
 36 Connelly, B.S. and Ones, D.S. 2008. *Interrater unreliability in assessment center ratings: A meta-analysis*. 36
 37 Paper presented at the Annual Conference of the Society for Industrial and Organizational 37
 38 Psychology, April, San Francisco, CA. 38
 39 Conway, J.M., Jako, R.A. and Goodman, D.F. 1995. A meta-analysis of interrater and internal 39
 40 consistency reliability of selection interviews. *Journal of Applied Psychology*, 80, 565–579. 40
 41 Conway, J.M. and Peneno, G.M. 1999. Comparing structured interview question types: Construct 41
 42 validity and applicant reactions. *Journal of Business and Psychology*, 13, 485–506. 42
 43 Dean, M.A., Roth, P.L. and Bobko, P. 2008. Ethnic and gender subgroup differences in assessment 43
 44 center ratings: a meta-analysis. *Journal of Applied Psychology*, 93, 685–691. 44
 45 Epstein, S. 1979. The stability of behavior: I. On predicting most of the people much of the time. 45
 46 *Journal of Personality and Social Psychology*, 37, 1097–1126. 46
 47 47

- 1 Hausknecht, J.P., Day, D.V. and Thomas, S.C. 2004. Applicant reactions to selection procedures: An
2 updated model and meta-analysis. *Personnel Psychology*, 57, 639–683. 2
- 3 Highhouse, S. and Harris, M.M. 1993. The measurement of assessment center situations: Bem's
4 template matching technique for examining exercise similarity. *Journal of Applied Social Psychology*, 4
5 23, 140–155. 5
- 6 Howard, A. 2008. Making assessment centers work the way they are supposed to. *Industrial and*
7 *Organizational Psychology*, 1, 98–104. 7
- 8 International Taskforce on Assessment Centers Guidelines. 2009. Guidelines and ethical
9 considerations for assessment center operations. *International Journal of Selection and Assessment*, 9
10 17, 243–253. 10
- 11 Irvine, S.H., Dann, P.L. and Anderson, J.D. 1990. Towards a theory of algorithm-determined
12 cognitive test construction. *British Journal of Psychology*, 81, 173–195. 12
- 13 Kudisch, J.D., Ladd, R.T. and Dobbins, G.H. 1997. New evidence on the construct validity of
14 diagnostic assessment centers: The findings may not be so troubling after all. *Journal of Social*
15 *Behavior and Personality*, 12, 129–144. 15
- 16 Lance, C.E. 2008. Why assessment centers do not work the way they are supposed to. *Industrial and*
17 *Organizational Psychology*, 1, 84–97. 17
- 18 Lievens, F. 2009, February. Effects of response fidelity on test performance and validity. Paper
19 presented at the 23rd Meeting of the Personnel and Human Resources Research Group, College
20 Station, TX. 20
- 21 Lievens, F. 2008. What does exercise-based assessment really mean? *Industrial and Organizational*
22 *Psychology: Perspectives on Science and Practice*, 1, 117–120. 22
- 23 Lievens, F. 2001. Assessor training strategies and their effects on accuracy, inter-rater reliability, and
24 discriminant validity. *Journal of Applied Psychology*, 86, 255–264. 24
- 25 Lievens, F. and Conway, J.M. 2001. Dimension and exercise variance in assessment center scores: A
26 large-scale evaluation of multitrait-multimethod studies. *Journal of Applied Psychology*, 86, 1202–
27 1222. 27
- 28 Lievens, F., Keen, G. and Schollaert, E. 2010, April. *A novel look at behavior elicitation in assessment*
29 *center exercises*. Poster session presented at the 25th Annual Conference of the Society for Industrial
30 and Organizational Psychology, Atlanta, United States. 30
- 31 Lievens, F. and Sackett, P.R. 2007. Situational judgment tests in high stakes settings: Issues and
32 strategies with generating alternate forms. *Journal of Applied Psychology*, 92, 1043–1055. 32
- 33 Lievens, F., Tett, R.P. and Schleicher, D.J. 2009. Assessment centers at the crossroads: Toward a
34 reconceptualization of assessment center exercises. In *Research in Personnel and Human Resources*
35 *Management*, edited by J.J. Martocchio and H. Liao. Bingley: JAI Press, 99–152. 35
- 36 McDaniel, M.A., Hartman, N.S., Whetzel, D.L. and Grubb, W.L. 2007. Situational judgment tests,
37 response instructions, and validity: A meta-analysis. *Personnel Psychology*, 60, 63–91. 37
- 38 McFarland, L.A., Yun, G.J., Harold, C.M., Viera, L. and Moore, L.G. 2005. An examination of
39 impression management use and effectiveness across assessment center exercises: The role of
40 competency demands. *Personnel Psychology*, 58, 949–980. 40
- 41 Meyer R.D., Dalal, R.S. and Hermida, R. 2010. A review and synthesis of situational strength in the
42 organizational sciences. *Journal of Management*, 36, 121–140. 42
- 43 Motowidlo, S.J., Hooper, A.C. and Jackson, H.L. 2006. Implicit policies about relations between
44 personality traits and behavioral effectiveness in situational judgment items. *Journal of Applied*
45 *Psychology*, 91, 749–761. 45
- 46 46
- 47 47

1	Reilly, R.R., Henry, S. and Smither, J.W. 1990. An examination of the effects of using behavior	1
2	checklists on the construct validity of assessment center dimensions. <i>Personnel Psychology</i> , 43,	2
3	71–84.	3
4	Rupp, D.E., Gibbons, A.M. and Snyder, L.A. 2008. Transforming our models of learning and	4
5	development: Web-based instruction as enabler of third-generation instruction. <i>Industrial and</i>	5
6	<i>Organizational Psychology: Perspectives on Science and Practice</i> , 1, 454–467.	6
7	Schneider, J.R. and Schmitt, N. 1992. An exercise design approach to understanding assessment-	7
8	center dimension and exercise constructs. <i>Journal of Applied Psychology</i> , 77, 32–41.	8
9	Schollaert, E. and Lievens, F. in press. The use of role-player prompts in assessment center exercises.	9
10	<i>International Journal of Selection and Assessment</i> .	10
11	Schollaert, E. and Lievens, F. 2009, October. <i>Using trait activation theory to increase the amount of</i>	11
12	<i>candidate behaviors elicited in assessment center exercises?</i> Paper session presented at the 4th Dutch-	12
13	Flemish Research Meeting on Personnel Recruitment and Selection, Amsterdam, The Netherlands.	13
14	Stricker, L.J. and Rock, D.A. 1990. Interpersonal competence, social intelligence, and general ability.	14
15	<i>Personality and Individual Differences</i> , 11, 833–839.	15
16	Tett, R.P. and Burnett, D.D. 2003. A personality trait-based interactionist model of job performance.	16
17	<i>Journal of Applied Psychology</i> , 88, 500–517.	17
18	Tett, R.P. and Guterman, H.A. 2000. Situation trait relevance, trait expression, and cross-situational	18
19	consistency: Testing a principle of trait activation. <i>Journal of Research in Personality</i> , 34, 397–423.	19
20	Thornton, G.C. III and Mueller-Hanson, R.A. 2004. <i>Developing Organizational Simulations: A Guide for</i>	20
21	<i>Practitioners and Students</i> . Mahwah, NJ: Lawrence Erlbaum Associates, Inc.	21
22		22
23		23
24		24
25		25
26		26
27		27
28		28
29		29
30		30
31		31
32		32
33		33
34		34
35		35
36		36
37		37
38		38
39		39
40		40
41		41
42		42
43		43
44		44
45		45
46		46
47		47