PERSONNEL ASSESSMENT AND DECISIONS

Personnel Assessment and Decisions

Volume 5 | Issue 1 Article 5

7-6-2019

A New Scoring Procedure in Assessment Centers: Insights from Interaction Analysis

Janneke K. Oostrom
Vrije Universiteit Amsterdam, the Netherlands, j.k.oostrom@vu.nl

Nale Lehmann-Willenbrock
University of Hamburg, Germany, nale.lehmann-willenbrock@uni-hamburg.de

Ute-Christine Klehe
Justus-Liebig-Universität Gießen, Germany, Ute-Christine.Klehe@psychol.uni-giessen.de

Recommended Citation
Oostrom, Janneke K.; Lehmann-Willenbrock, Nale; and Klehe, Ute-Christine (2019) "A New Scoring Procedure in Assessment Centers: Insights from Interaction Analysis," *Personnel Assessment and Decisions*: Vol. 5 : Iss. 1 , Article 5.
DOI: 10.25035/pad.2019.01.005
Available at: https://scholarworks.bgsu.edu/pad/vol5/iss1/5

A New Scoring Procedure in Assessment Centers: Insights From Interaction Analysis

Janneke K. Oostrom[1], Nale Lehmann-Willenbrock[2], and Ute-Christine Klehe[3]

1. Vrije Universiteit Amsterdam, the Netherlands
2. University of Hamburg, Germany
3. Justus-Liebig-Universität Gießen, Germany

KEYWORDS

assessment centers, interaction analysis, personnel selection

ABSTRACT

This paper proposes interaction analysis as an alternative scoring procedure in assessment centers (ACs). Interaction analysis allows for a more fine-grained scoring approach by which candidate behaviors are captured as they actually happen, thus avoiding judgment errors typically associated with traditional scoring procedures. We describe interaction analysis and explain how this procedure can improve the validity of ACs. In a short research example, we showcase how interaction analysis can be implemented in AC settings. Finally, we integrate our arguments in terms of three key propositions that we hope will inspire future research on more dynamic scoring procedures.

ACs are widely applied around the globe, most frequently to managerial jobs ranging from supervisor to executive (Spychalski, Quiñones, Gaugler, & Pohley, 1997). ACs are a popular assessment instrument because of their predictive validity (Becker, Höft, Holzenkamp, & Spinath, 2011; Meriac, Hoffman, Woehr, & Fleisher, 2008) and favorable candidate reactions in comparison with other assessment instruments (Hausknecht, Day, & Thomas, 2004). ACs typically last one or two days (Krause & Thornton, 2009) and consist of various simulation exercises (e.g., interviews, role-plays, presentations, in-basket exercises, and group discussions) that are evaluated by multiple trained assessors on multiple job-related dimensions (Lievens & Thornton, 2005).

Despite their widespread use, current ACs are often criticized (e.g., Highhouse, 2002; Kuncel, Klieger, Connelly, & Ones, 2013). For instance, several meta-analyses have shown that ACs still perform worse in predicting performance than simple cognitive ability tests (e.g., Schmidt & Hunter, 1998). One explanation for this recurrent finding is that ACs highly depend on assessors' subjective ratings of candidates' behaviors. Several authors have therefore highlighted the need for alternative scoring procedures to make more accurate AC judgments (e.g., Silzer & Jeanneret, 2011). As one potential avenue to address this need, we propose interaction analysis as an alternative scoring procedure for all AC exercises that involve actual interactions between two or more individuals (e.g., interviews, role plays, presentations, and group discussions). Through interaction analysis candidate behaviors are captured as they actually happen, thereby avoiding judgment errors typically associated with traditional scoring procedures. The purpose of this paper is to explain the basic steps in interaction analysis and to showcase how this scoring procedure can be implemented in ACs. The paper ends with three key propositions regarding the predictive validity, construct validity, and acceptability of ACs using interaction analysis.

Interaction Analysis in ACs

Interaction analysis is a methodological approach that has been applied across a broad range of research domains. Examples are diverse and include studies on change management (e.g., Klonek, Lehmann-Willenbrock, & Kauffeld, 2014) and leader –follower dynamics (Lehmann-Willenbrock, Meinecke, Rowold, & Kauffeld, 2015). Yet, to the best of our knowledge, personnel selection research has not explored the possibilities of fine-grained interaction analyses to date. Although this new approach still needs assessors to evaluate candidate behaviors (like in traditional

Corresponding author:
Janneke K. Oostrom
Email: j.k.oostrom@vu.nl

ACs), the codings of the behaviors are far less subjective; behavioral units are classified and not immediately judged on their effectiveness, nor linked to a trait or competency, thereby reducing the risks of rater errors. Therefore, we think interaction analysis will result in more accurate and less subjective evaluations than traditional AC scoring procedures. Table 1 provides an overview of the differences between the traditional AC scoring procedure and the new procedure.

Basic Steps in Interaction Analysis

Although specific research questions and applications across these different settings differ widely, the general approach to understanding and analyzing behavioral processes using interaction analysis is quite similar. The following basic steps have been described in detail in Lehmann-Willenbrock and Allen (2018) as well as Meinecke and Lehmann-Willenbrock (2015). Here, we apply them to the specific case of AC exercises. First, the interested researcher will need to set up the behavioral data gathering. To this end, most previous studies using a quantitative interaction analytical approach rely on videotaped behavioral data, which allows the identification of both verbal and nonverbal behavior. It can be played back repeatedly for additional or follow-up analyses and can also be used for training and feedback material at a later point. Previous research suggests that groups tend to ignore or forget the camera as soon as a group discussion is under way (Kauffeld & Lehmann-Willenbrock, 2012). As long as only verbal behavior is of interest or when videotaping is not possible, audiotaped data may also be an option (e.g., Meinecke, Lehmann-Willenbrock, & Kauffeld, 2017).

Once video (or, less ideally, audio) data become accessible, the specific phenomena that are to be identified from the data have to be defined. Subject matter experts could be asked to develop coding schemes for specific AC dimensions (e.g., integrity, valuing diversity, adaptability, problem solving, or conflict resolution). However, using an existing coding scheme is often preferable, as findings can then be related to theoretical models. Coding schemes generally focus on the occurrences of specific behaviors: coders label each specific behavior taking place during the exercise (e.g., "suggesting a solution" or "presenting an idea") without making inferences about the candidate's traits or competencies (see Table 2 for an example of a coding scheme). Here, the difference between the traditional AC scoring procedure and interaction analysis becomes clear. If, for example, the goal of an AC exercise is to assess how candidates approach and solve a complex problem, assessors would traditionally use some kind of Likert-type scale to score the candidate's overall skills based on the behaviors they have seen during the exercise. With interaction analysis, the occurrence of any specific behavior related to problem solving (e.g., "describing a problem," "defining the objective," or "describing a solution") is coded as it actually happened in time. Upon deciding on a coding scheme that is suitable for analyzing the relevant question(s) and capturing the behavioral units of interest, coders need to be trained in order to establish inter-rater reliability. In the case of interaction analysis, inter-rater reliability is examined by having several (i.e., at least two) trained raters code the same video material and calculating the degree to which they reach the same conclusions regarding each coded behavioral unit.

Then, the question of unitizing needs to be addressed: Where does one behavioral unit start and stop, and when will a new unit be assigned? Unitizing rules can differ depending on the goal of the assessment (Meinecke & Lehmann-Willenbrock, 2015) but typically adhere to one of the following rules: (a) turns of talk (i.e., assign a new behavioral unit as soon as the speaker changes; e.g., Chiu, 2008), (b) utterances (i.e., assign a new behavioral unit when a functionally different statement begins; e.g., Lehmann-Willenbrock et al., 2015), or (c) specific temporal segments within a conversation (e.g., 2-minute segments, Barsade, 2002; or predefined group discussion phases). Within ACs, we would recommend either unitizing based on turns of talk or utterances, as this would allow focusing on candidate behaviors at the individual level. Note that a unitizing rule based on specific utterances could mean that two consecutive behavioral units are contributed by the same candidate, for example when a candidate voices an idea and immediately follows up with a question to the other candidates. Unitizing based on specific temporal segments could be useful when one is interested in answering more generalized research questions, such as how specific behaviors (e.g., humor) within certain time fragments (e.g., the start of the exercise) affect overall assessment ratings.

Once inter-rater reliability is established and the data have been coded, there are several options for examining the annotated data. These include frequency analysis, co-occurrence analysis, lag sequential analysis, or pattern analysis for identifying behavioral triggers and emergent behavioral patterns (for an overview, see Meinecke & Lehmann-Willenbrock, 2015). There are several software solutions available that facilitate the coding process substantially, such that traditional transcripts of the observed behaviors are no longer required. These software solutions preserve the temporal order of the interaction data by registering time stamps (i.e., onset and offset times) along with each coded behavior (for an overview and comparison of possible software options, see Lehmann-Willenbrock and Allen, 2018).

TABLE 1.

Overview of Differences Between the Traditional AC Evaluation Method and Interaction Analysis

	Traditional AC scoring procedure	Interaction analysis	Potential benefits of interaction analysis
Preparation	Conducting a job analysis to determine relevant competencies and traits	Choosing/developing a coding scheme of relevant behaviors	More evidence for validity of existing coding schemes
	Choosing/developing a simulation exercise that will elicit relevant trait-driven behaviors	Choosing/developing a simulation exercise that will elicit those behaviors	Easier to elicit behavioral utterances than trait-driven behaviors
	Training the assessors in the rating process	Training the assessors in the rating process	None
Procedure	Observing the candidate(s) during the simulation exercise	Recording the simulation exercise	Assessor(s) do(es) not need to be present
			Recordings remain available for later in-depth analysis and follow-up
	Discussing observations among assessors	Unitizing of recorded material and coding of actual behaviors based on recordings (typically supported by software)	Codings are less subjective and less prone to rater errors
			Less potential for persuasion among assessors
	Scoring traits and/or competencies immediately after the exercise	Analyzing the data, e.g., lag sequential analysis or pattern analysis for identifying behavioral triggers and emergent temporal patterns	Rich information which can be analyzed in different ways
Insights	Trait-related dimensions	Specific behaviors and processes	Insights into unfolding interaction patterns and behaviors that are activated in realistic job situations
		Dynamic interaction patterns and context effects	More in depth-information about individual candidates' behaviors as well as interdependencies between several candidates within the same exercise

Applying Interaction Analysis in ACs: An Example

To illustrate how interaction analysis can inform and improve ACs, we showcase the results of a laboratory study. In this study, we set up videotaped group discussions that exemplify the typical group setup in an AC exercise. These group discussions were videotaped, and an interaction analytical procedure was used in order to explore the utility of this method for measuring leadership. Note that although we applied interaction analysis to a leaderless group discussion, this new scoring procedure can be used for any type of AC exercise that involves actual interactions (e.g., interviews and role plays).

TABLE 2.

Act4teams Coding Scheme for Verbal Behavior During Group Interactions (e.g., Kauffeld & Lehmann-Willenbrock, 2012)

Problem-focused behaviors	Procedural behaviors	Socio-emotional behaviors	Action-oriented behaviors
Endorsing a problem	**Positive procedural behaviors:**	**Positive socio-emotional behaviors:**	**Positive, proactive behaviors:**
Problem	Goal orientation	Encouraging participation	Expressing positivity
Describing a problem	Clarifying	Providing support	Taking responsibility
Connections with problems	Procedural suggestion	Active listening	Action planning
Defining the objective	Procedural question	Reasoned disagreement	
Solution	Prioritizing	Giving feedback	**Negative, counterproductive behaviors:**
Describing a solution	Time management	Humor	No interest in change
Problem with a solution	Task distribution	Laughter	Complaining
Arguing for a solution	Visualization	Separating opinions from facts	Denying responsibility
Organizational knowledge	Weighting costs and benefits	Expressing feelings	Empty talk
Knowing who	Summarizing	Offering praise	Ending the discussion early
Question			
	Negative procedural behaviors:	**Negative socio-emotional behaviors:**	
	Losing the train of thought (running off topic)	Criticizing/backbiting	
		Interrupting	
		Side conversations	
		Self-promotion	

Sample and Procedure

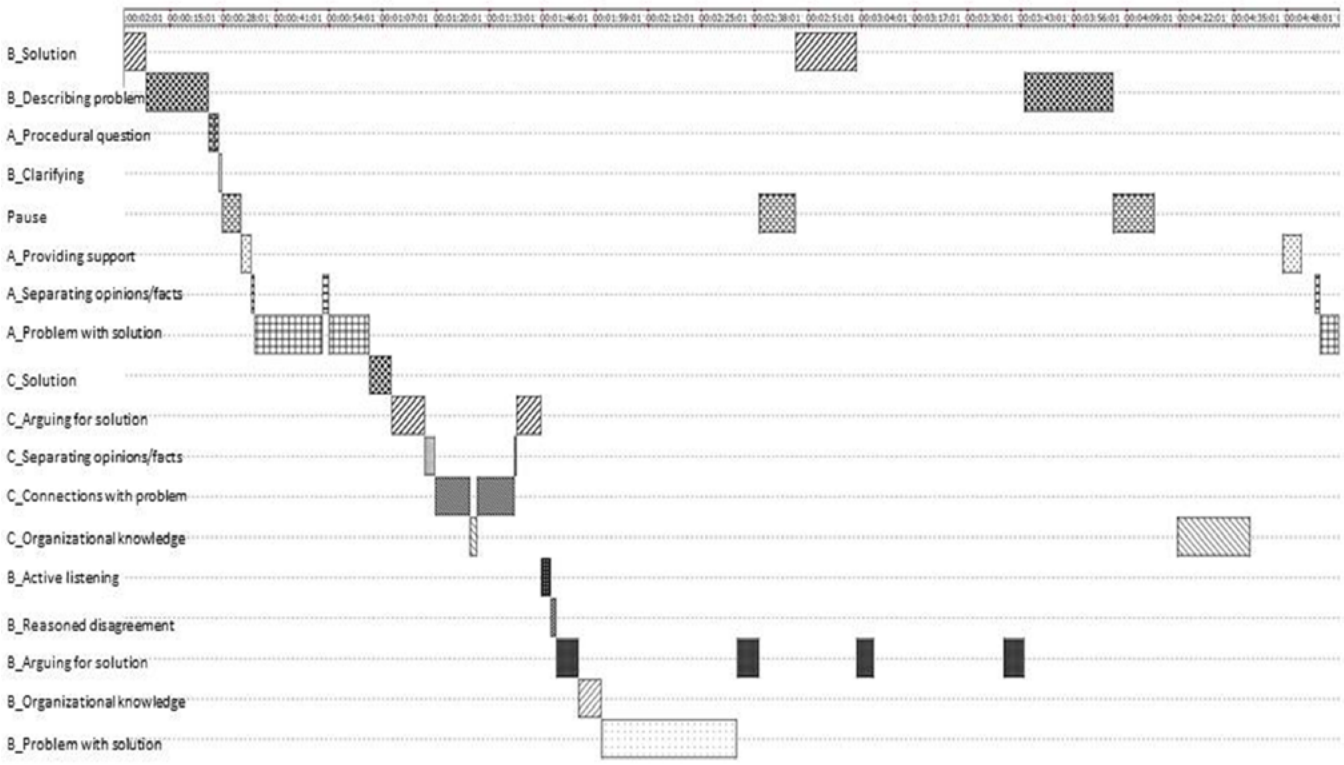
We recruited 30 groups of three participants at a large university in the Netherlands. The majority of the participants were psychology students, and two-thirds of them were female (60 out of 90 participants). Their age ranged from 18 to 34 years (M = 22.64, SD = 3.67). Participants could choose from earning participation credits or 10 euros of remuneration. The experiment was formally approved by the ethics committee at the participating university. Each participant was randomly assigned to one of three roles in a leaderless group discussion (i.e., HR manager, production manager, or sales manager) and provided with unique role-based background information that needed to be revealed and synthesized to reach a solution (Klehe et al., 2012; Klehe, König, Richter, Kleinmann, & Melchers, 2008). Participants were given 10 minutes to prepare and up to 30 minutes for the actual group discussion.

Measures

The discussions were rated by two randomly chosen assessors out of a pool of five trained graduate students, using traditional AC observer rating sheets (see Appendix, in this study we focused on the leadership items; α = .82; ICC = .78). In addition, we videotaped all interactions and analyzed the data using the act4teams coding scheme (Table 2) implemented in INTERACT software (Mangold, 2010). The videos were independently coded by two trained graduate students. As depicted in Table 2, the act4teams scheme describes (verbal) social interactions in terms of four broader dimensions: problem-focused behaviors, procedural behaviors, socio-emotional behaviors, and action-oriented behaviors. Although this coding scheme has been developed to score a broad range of behaviors in team contexts and not necessarily emergent leadership behaviors, there is considerable overlap between the behaviors in the act4teams coding scheme and emergent leadership behaviors (e.g., Kickul & Neuman, 2000; Lord, Phillips, & Rush, 1980). For details on the theoretical background of this coding scheme and its development and validation, see Kauffeld and Lehmann-Willenbrock (2012). In order to establish the reliability of our coding approach, five interactions were coded by both students, showing sufficient inter-rater agreement (Cohen's κ =.75).

Figure 1.

Sample segment (first 5 minutes) showing participants and specific verbal contents. Each line in the graph represents one specific behavior by one specific participant. For example, the first lines shows instances (and time stamps) when participant B contributed solutions in the discussion process.



Individual Behaviors Related to Leadership Potential

To explore which behavioral patterns were indicative of leadership, we related the each specific type of behavior coded with the act4teams scheme to the overall leadership score obtained from the traditional AC rating for each individual in the group. To do so, we enumerated the absolute frequency of each specific type of behavior coded with the act4teams scheme per observed individual participant. The discussion varied somewhat in duration (M = 19.77 minutes; range = 11–30), and so we summed the absolute frequency of each behavioral category per participant and related this frequency to a 20-minute period (i.e., dividing each of them by the respective discussion length and multiplying by 20). We then calculated Pearson's correlations at the individual level (N = 90) to explore the relationship between the coded behaviors and the overall leadership rating for each participant.

Specific verbal behaviors observed during the group discussion were meaningfully linked to the overall rating of a candidate's leadership score. These behaviors largely presented procedural behaviors aimed at structuring the discussion (goal orientation, r = .32, p < .01; clarifying, r = .31, p < .01; procedural suggestion, r = .37, p < .01; procedural question, r = .29, p < .01; time management, r = .24, p < .01; summarizing, r = .31, p < .01), but also select problem-focused behaviors (describing a solution, r = .21, p = .05; organizational knowledge, r = .23, p = .03; knowing who, r = .26, p = .02; question, r = .35, p < .01), and select positive socio-emotional behaviors (giving feedback, r = .37, p < .01; use of humor, r = .21, p = .05). Taken together, these findings suggest that AC assessors mainly react to expressions of procedural behaviors when making leadership ratings.

When examining individual items in the leadership rating on the traditional observation sheet (Appendix), the additional findings obtained from interaction analysis yield more specific insights into otherwise relatively vague descriptions of leadership. For example, the item L2 in Appendix ("manages the discussion") does not specify how this is actually accomplished by the candidate. A closer inspection of the behavioral correlates of this item highlights procedural behaviors such as goal orientation (r = .35, p < .01), clarifying (r = .28, p < .01), procedural suggestions (r = .38, p < .01), procedural questions (r = .29, p < .01), time management (r = .32, p < .01), task distribution (r = .30, p < .01), and summarizing (r = .39, p < .01).

Graphing the Interaction Process

Figure 1 illustrates how these different behaviors unfolded over the course of one exemplary AC group discussion. Using INTERACT software, we can "zoom in" on particular discussion segments. For illustrative purposes, Figure 1 only shows the first 5-minute segment from this group's discussion. The top of Figure 1 shows the actual time line from this discussion. Each consecutive line shows specific behaviors by each of the participants in this group discussion. The discussion was initiated by participant B, who contributed a solution. Then B proceeded to describe a problem, followed by a procedural question by participant A, and so forth. Graphic depictions of the fine-grained details of the discussion process are helpful for exploring the communication dynamics during such AC exercises, for understanding the role of individual participants within the social context, and for identifying ("eyeballing") potentially critical statements or behavioral triggers, which can then be followed up by in-depth quantitative analyses.

Identifying Interaction Patterns

Whereas traditional ratings in ACs focus on the individual only, interaction analysis also allows us to consider sequences or patterns of behavior. Lag sequential analysis can test how specific behaviors by candidates during the AC exercise trigger other behaviors. Significant behavior patterns are identified by z-values larger than 1.96. In our research example presented here, at lag1 (i.e., behavior sequences from one behavior immediately to the next) we found that support by other group members was triggered by goal orientation (z = 13.68), by clarifying (z = 311.97), by task distribution (z = 3.18), time management (z = 5.69), and summarizing (z = 24.13). Hence, those procedural behaviors that were linked to overall leadership ratings by observers in fact also triggered support by other group members within the interaction process.

Rater Errors and Gender Stereotypes

The current data allowed us to test both the occurrence of a halo effect and the use of gender stereotypes when contrasting the two different scoring procedures. For the traditional observation sheet (Appendix), the mean intercorrelation of .59 between the four dimensions (i.e., planning, cooperation, leadership, and communication) suggests a pervasive halo error. In contrast, the codings of the specific behaviors did not show such a halo effect: The mean absolute intercorrelation of the codings was only .11. At the overall dimension level, the mean intercorrelation was .34, which is still considerably lower than the intercorrelation of the traditional scoring procedure.

Furthermore, the traditional leadership rating showed a substantial score difference in favor of male candidates (t[88] = 3.56, p < .01, d = 0.78; M = 3.73, SD = 0.55 for men; M = 3.31, SD = 0.53 for women). In comparison, the behavioral codings did not show a clear gender bias. We found gender differences for only four specific behaviors, three in favor of men (summarizing, t[88] = 2.48, p = .02, d = 0.53; M = 0.91, SD = 1.42 for men; M = 0.32, SD = 0.82 for women; giving feedback, t[88] = 2.04, p = .05, d = 0.43; M = 2.65, SD = 2.33 for men; M = 1.68, SD = 2.00 for women; criticizing others, t[88] = -2.29, p = .02, d = -0.49; M = 0.17, SD = 0.49 for men; M = 1.00, SD = 1.94 for women) and one in favor of women (side conversations, t[88] = 2.22, p = .03, d = 0.47; M = 2.04, SD = 2.39 for men; M = 1.05, SD = 1.79 for women).

Future Research Agenda

ACs highly depend on assessors' subjective ratings of candidates' behaviors. Several authors have therefore highlighted the need to move away from "gut feelings" and subjective ratings and toward a more fine-grained and objective scoring procedure (e.g., Silzer & Jeanneret, 2011). We have proposed an alternative scoring procedure in ACs: interaction analysis. Through interaction analysis candidate behaviors are captured as they actually happen, thereby avoiding judgment errors typically associated with traditional scoring procedures. In this paragraph we discuss the validity and acceptability of this alternative approach and integrate our arguments in terms of three key propositions.

Predictive Validity

The most important difference between traditional AC scoring procedures and interaction analysis is that instead of relying on overall rater observations of behavior, specific behavioral observations are used to predict performance. These behavioral observations address the social context in which each behavior occurs by studying its direct antecedents and consequences. For example, in a leaderless group discussion, interaction analysis can show how specific behaviors by candidates during the AC exercise trigger other candidates' behaviors (or, in case of interviews or role plays, the interviewer's or actor's behaviors). This is relevant as not every conceptually useful behavior will be useful at every point in time; the same behavior may be much more useful at the start of an exercise than at the end when all information has possibly already been shared and discussed. Such a differentiation is usually not considered in traditional observation sheets but can well be taken into account in interaction analyses. Thus, group discussion-based AC exercises and the intricate social dynamics inherent in them the complexity of interpersonal relations and unfolding interaction patterns that characterize real job situations (e.g., Lehmann-Willenbrock & Allen, 2018). For this reason, we expect interaction analysis to allow for a stronger predictor-criterion alignment, and hence more predictive power of the AC (Arthur & Villado, 2008). Furthermore, a focus on actual behavioral expressions embedded in social

interactions instead of more abstract traits and competencies might be beneficial for the predictive and incremental validity of ACs, as traits and competences might be more economically and objectively captured via personality questionnaires and cognitive ability tests (Meriac et al., 2008). Based on these arguments, we formulated the following proposition.

Proposition 1: ACs using behavioral ratings derived from interaction analyses have higher predictive validity than ACs using traditional scoring procedures.

We believe the predictive validity of ACs to especially benefit from using interaction analysis when predicting behavioral criteria (e.g., interpersonal or communication skills, decision-making, citizenship behaviors), as these allow for the strongest predictor-criterion alignment.

Construct Validity

To date, the construct validity of ACs remains somewhat elusive because different dimensions within exercises correlate higher than similar dimensions across exercises (e.g., Wirz, Melchers, Schultheiss, & Kleinmann, 2014; Woehr & Arthur, 2003). Several studies have already demonstrated increases in the construct validity of ACs by improving the observation of the behaviors shown during the exercise. This was accomplished either by reducing the number of dimensions to rate (Kolk, Born, & Van der Flier, 2004), by ensuring that the behaviors to be rated will be visible in the exercise (Klehe et al., 2008; Lievens, Chasteen, Day, & Christianson, 2006), by frame-of-reference trainings (Woehr & Huffcutt, 1994), and by using behavioral checklists instead of overall trait ratings (Jackson, Barney, Stillman, & Kirkley, 2007). Although the effects on construct validity tend to be small, these findings suggest that more systematic procedures that enable AC developers to select independent and easily measurable (behavioral) dimensions will help distinguish between these dimensions. Interaction analysis goes one step further than these previously suggested methods as it allows for differentiation based on identifiable and differentiable behaviors as they happen during the AC exercise. In addition, by using interaction analysis the raters can focus on behaviors of interest that can be observed independent of the exercises. For example, behavioral checklists are completed by assessors immediately after an exercise. Behavioral checklist are therefore more cognitively demanding than interaction analysis, as the assessor has to observe, recognize, and recall the behaviors of each of the candidates (Reilly, Henry, & Smither, 1990). In contrast, interaction analysis makes use of recordings that can be played back as often as needed. Interaction analysis also provides additional advantages over frame-of-reference training. Although both methods can reduce rater errors (including the halo effect), a frame-of-reference training does not guarantee changes in assessor behaviors, nor does it make the evaluation procedure less subjective. Interaction analysis, however, forces the assessor to focus on the actual behaviors that are being demonstrated during the exercise. Based on these arguments, we formulated the following proposition.

Proposition 2: Interaction analysis improves the construct validity of AC exercises.

Acceptability

In order for interaction analysis to be a viable measurement approach in ACs and to be accepted by assessors and candidates, the benefits should outweigh the potential costs. Compared to other selection instruments, an AC is already an expensive, complex, and labor intensive procedure. Interaction analysis requires videotaping exercises and coding the behaviors of each candidate, which makes ACs potentially an even more time consuming and expensive procedure. However, these costs might be reduced in the near future, as modern technology such as latent semantic analysis (e.g., Campion, Campion, Campion, & Reider, 2016) and social sensing technology (Schmid Mast, Gatica-Perez, Frauendorfer, Nguyen, & Choudhury, 2015) might allow for automatic scoring of behaviors.

Typically, candidates receive some initial feedback at the end of day. Using interaction analysis would reduce the speed at which any feedback to candidates can be provided. Furthermore, not every candidate (especially candidates with high-level executive roles or candidates that have been headhunted) may allow videotaping or any other proof of their interest in other jobs until selection decisions are final. For these reasons, we expect there will be some limitations to the application of the interaction approach. However, we do think that such an approach can be used in a broad range of ACs, including ACs that are used in development, coaching, or promotion programs. We believe that in such programs, an interaction approach has a number of other benefits for both assessors and candidates. First of all, interaction analyses extract more information from the same interaction sequences than traditional AC observations. As such, they can generate more in-depth information while using existing AC exercises (i.e., in this case no additional investment costs are incurred for developing new exercises). A more reliable and valid approach, based on such in-depth behavioral observations, has benefits for each party involved. Second, because ACs are often used for developmental purposes rather than selection purposes (Hazucha et al., 2011), an in-depth analysis of a candidate's behaviors and others' responses to those behaviors (i.e., temporal patterns) can offer promising practical implications for development purposes. Showing participants their own behaviors

and helping them reflect about how those behaviors helped them succeed within the social process is likely to be more informative than the feedback from more traditional ratings as currently used in AC practice. For these reasons, we believe that —when it is possible to videotape AC exercises and use interaction analyses—the benefits of this approach outweigh the potential costs for both assessors and candidates.

Proposition 3: The acceptability of using interaction analyses, especially the type of feedback it provides, is higher than the acceptability of traditional AC scoring procedures

Conclusion

The purpose of this paper is to draw attention to interaction analysis as an alternative scoring procedure in ACs and to showcase how this scoring procedure can be implemented in ACs. We have integrated our arguments in terms of three key propositions regarding the validity and acceptability of ACs using interaction analysis, which we hope will inspire future research.

REFERENCES

Arthur Jr, W., & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. Journal of Applied Psychology, 93, 435-442. doi:10.1037/0021-9010.93.2.435

Barsade, S. G. (2002). The ripple effect: Emotional contagion and its influence on group behavior. Administrative Science Quarterly, 47, 644-675. doi:10.2307/3094912

Becker, N., Höft, S., Holzenkamp, M., & Spinath, F. M. (2011). The predictive validity of assessment centers in German-speaking regions. Journal of Personnel Psychology, 10, 61-69. doi:10.1027/1866-5888/a000031

Campion, M. C., Campion, M. A., Campion, E. D., & Reider, M. H. (2016). Initial investigation into computer scoring of candidate essays for personnel selection. Journal of Applied Psychology, 101, 958-975. doi:10.1037/apl0000108

Chiu, M. M. (2008). Flowing toward correct contributions during group problem solving: A statistical discourse analysis. Journal of the Learning Sciences, 17, 415-463. doi:10.1080/10508400802224830

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. doi:10.1111/j.1744-6570.2004.00003.x

Hazucha, J. F., Ramesh, A., Goff, M., Crandell, S., Gerstner, C., Sloan, E., ... & Van Katwyk, P. (2011). Individual psychological assessment: The poster child of blended science and practice. Industrial and Organizational Psychology, 4, 297-301. doi:10.1111/j.1754-9434.2011.01342.x

Highhouse, S. (2002). Assessing the candidate as a whole: A historical and critical analysis of individual psychological assessment for personnel decision making. Personnel Psychology, 55, 363-396. doi:10.1111/j.1744-6570.2002.tb00114.x

Jackson, D. J., Barney, A. R., Stillman, J. A., & Kirkley, W. (2007). When traits are behaviors: The relationship between behavioral responses and trait-based overall assessment center ratings. Human Performance, 20, 415-432. doi:10.1080/08959280701522130

Kauffeld, S., & Lehmann-Willenbrock, N. (2012). Meetings matter: Effects of team meetings on team and organizational success. Small Group Research, 43, 130–158. doi:10.1177/1046496411429599

Kickul, J., & Neuman, G. (2000). Emergent leadership behaviors: The function of personality and cognitive ability in determining teamwork performance and KSAs. Journal of Business and Psychology, 15, 27-51. doi:10.1023/A:1007714801558

Klehe, U.-C., Kleinmann, M., Hartstein, T., Melchers, K.G., König, C.J., Heslin, P., & Lievens, F. (2012). Responding to personality tests in a selection context: The role of the ability to identify criteria and the ideal-employee factor. Human Performance, 25, 273-302.

Klehe, U. C., König, C. J., Richter, G. M., Kleinmann, M., & Melchers, K. G. (2008). Transparency in structured interviews: Consequences for construct and criterion-related validity. Human Performance, 21, 107-137. doi:0.1080/08959280801917636

Klonek, F. E., Lehmann-Willenbrock, N., & Kauffeld, S. (2014). Dynamics of resistance to change: A sequential analysis of change agents in action. Journal of Change Management, 14, 334-360. doi:10.1080/14697017.2014.896392

Kolk, N. J., Born, M. Ph., & Van der Flier, H. (2004). A triadic approach to the construct validity of the assessment center: The effect of categorizing dimensions into a feeling, thinking, and power taxonomy. European Journal of Psychological Assessment, 20, 149-156. doi:10.1027/1015-5759.20.3.149

Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. Journal of Applied Psychology, 98, 1060-1072. doi:10.1037/a0034156

Krause, D. E., & Thornton III, G. C. (2009). A cross-cultural look at assessment center practices: Survey results from Western Europe and North America. Applied Psychology: An International Review, 58, 557-585. doi:10.1111/j.1464-0597.2008.00371.x

Lehmann-Willenbrock, N., & Allen, J. A. (2018). Modeling temporal interaction dynamics in organizational settings. Journal of Business and Psychology, 33, 325-344. doi:10.1007/s10869-017-9506-9

Lehmann-Willenbrock, N., Meinecke, A. L., Rowold, J., & Kauffeld,

S. (2015). How transformational leadership works during team interactions: A behavioral process analysis. Leadership Quarterly, 26, 1017-1033. doi:10.1016/j.leaqua.2015.07.003

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. doi:10.1037/0021-9010.91.2.247

Lievens, F., & Thornton, G. C. III. (2005). Assessment centers: recent developments in practice and research. In A. Evers, O. Smit-Voskuijl, & N. Anderson (Eds.), Handbook of selection (pp. 243-264). Malden, MA: Blackwell.

Lord, R. G., Phillips, J. S., & Rush, M. C. (1980). Effects of sex and personality on perceptions of emergent leadership, influence, and social power. Journal of Applied Psychology, 65, 176-182. doi:10.1037/0021-9010.65.2.176

Mangold. (2010). INTERACT quick start manual V2.4 (Mangold International GmbH, Ed.). Retrieved from www.mangold-international.com

Meinecke, A. L., & Lehmann-Willenbrock, N. (2015). Social dynamics at work: Meetings as a gateway. In J. A. Allen, N. Lehmann-Willenbrock & S. G. Rogelberg (Eds.), The Cambridge handbook of meeting science (pp. 325-356). New York, NY: Cambridge University Press.

Meinecke, A. L., Lehmann-Willenbrock, N., & Kauffeld, S. (2017). What happens during annual appraisal interviews? How leader-follower interactions unfold and impact interview outcomes. Journal of Applied Psychology, 102, 1054-1074. doi:10.1037/apl0000219

Meriac, J. P., Hoffman, B. J., Woehr, D. J., & Fleisher, M. S. (2008). Further evidence for the validity of assessment center dimensions: A meta-analysis of the incremental criterion-related validity of dimension ratings. Journal of Applied Psychology, 93, 1042-1052. doi:10.1037/0021-9010.93.5.1042

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. doi:10.1111/j.1744-6570.1990.tb02006.x

Schmid Mast, M., Gatica-Perez, D., Frauendorfer, D., Nguyen, L., & Choudhury, T. (2015). Social sensing for psychology: Automated interpersonal behavior assessment. Current Directions in Psychological Science, 24, 154-160. doi:10.1177/0963721414560811

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. Psychological Bulletin, 124, 262-274. doi:10.1037/0033-2909.124.2.262

Silzer, R., & Jeanneret, R. (2011). Individual psychological assessment: A practice and science in search of common ground. Industrial and Organizational Psychology, 4, 270-296. doi:10.1111/j.1754-9434.2011.01341.x

Spychalski, A. C., Quiñones, M. A., Gaugler, B. B., & Pohley, K. (1997). A survey of assessment center practices in organizations in the United States. Personnel Psychology, 50, 71-90. doi:10.1111/j.1744-6570.1997.tb00901.x

Wirz, A., Melchers, K. G., Schultheiss, S., & Kleinmann, M. (2014). Are improvements in assessment center construct-related validity paralleled by improvements in criterion-related validity? Journal of Personnel Psychology, 13, 184-193. doi:10.1027/1866-5888/a000115

Woehr, D. J., & Arthur, W. (2003). The construct-related validity of assessment center ratings: A review and meta-analysis of the role of methodological factors. Journal of Management, 29, 231-258. doi:10.1177/014920630302900206

Woehr, D. J. & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. Journal of Occupational and Organisational Psychology, 67, 189-205. doi:10.1111/j.2044-8325.1994.tb00562.x

RECEIVED 04/04/18 ACCEPTED 04/22/19

Appendix
Observer Rating Form

Name of rater:_____

Date/session:_____

Participant ___(m/f) / _____		Desirable behavior	Undesirable behavior
Planning	Overall rating ☐	P 1 Evaluates everyone's facts and possibilities before making a decision P 2 Searches and asks for all available information P 3 Looks for the core of the problem P 4 Quickly structures complex situations/problems P 5 Considers alternatives P 6 Contributes ideas/ solutions	P 1 Ignores some possible alternatives P 2 Makes decisions too quickly P 3 Does not grasp the core of the problem P 4 Cannot structure complex situations/problems P 5 Does not see alternative possibi-lities/solutions P 6 Does not offer a solution
Cooperation	Overall rating ☐	C 1 Makes compromises C 2 Tries to mediate between different parties/opinions C 3 Picks up others' ideas, brings others into the discussion C 4 Actively includes others in the discussion C 5 Creates "win-win" situations C 6 Treats others with fairness and respect	C 1 Shows little will to compromise C 2 Blocks others' contributions, interrupts others C 3 Only considers own ideas and solutions C 4 Talks in monologues C 5 Only considers own goals C 6 Treats others unfairly
Leadership	Overall rating ☐	L 1 Takes initiative L 2 Manages the discussion L 3 Checks processes and results L 4 Represents and defends own opinion/position L 5 Distributes tasks, delegates L 6 Gets others' attention; others follow his/her ideas	L 1 Lets others ask questions without showning initiative him-/herself L 2 Does not manage the discussion L 3 Does not check processes and results L 4 Does not defend own opinion/position L 5 Accepts tasks from others/follows L 6 Subordinates him-/herself
Communication	Overall rating ☐	C 1 Fluid,comprehensible expressions C 2 Uses adequate gestures and mimics C 3 Uses concrete, comprehensible arguments C 4 Provides other with relevant information C 5 Makes an effort to get everyone to an adequate state of knowledge C 6 Listens and lets others finish	C 1 Stagnant, hesistant expressions C 2 Gestures and mimics seem out of place C 3 Uses incomprehensible arguments C 4 Holds back relevant information C 5 Does not show interest in others' state of knowledge C 6 Cuts other participants off (interrupts)
	Rating scale:	**5** 4.5 **4** 3.5 **3** ☺ 😐	2.5 **2** 1.5 **1** ☹