Industrial and Organizational Psychology (2024), **17**, 192–205
doi:10.1017/iop.2024.2

FOCAL ARTICLE

Any slice is predictive? On the consistency of impressions from the beginning, middle, and end of assessment center exercises and their relation to performance

Pia V. Ingold[1], Anna Luca Heimann[2], and Simon M. Breil[3]

[1]Department of Psychology, University of Copenhagen, Denmark, [2]Department of Psychology, University of Zurich, Switzerland and [3]Department of Psychology, University of Münster, Germany
Corresponding author: Pia V. Ingold; Email: pia.ingold@psy.ku.dk

(Received 19 December 2023; accepted 3 January 2024; first published online 18 March 2024)

Abstract
This study generates new insights on the role of initial impressions in assessment centers. Drawing from the "thin slices" of behavior paradigm in personality and social psychology, we investigate to what extent initial impressions of assessees—based on different slices of assessment center exercises (i.e., two minutes at the beginning, middle, and end of AC exercises)—are consistent across and within AC exercises, and are relevant for predicting assessment center performance and job performance. Employed individuals ($N = 223$) participated in three interactive assessment center exercises, while being observed and evaluated by trained assessors. Based upon video-recordings of all assessment center exercises, a different, untrained group of raters subsequently provided ratings of their general initial impressions of assessees for the beginning, middle, and end of each exercise. As criterion measure, supervisors rated assessees' job performance. Results show that initial impressions in assessment centers are (a) relatively stable, (b) consistently predict assessment center performance across different slices of behavior (i.e., across the three time points and exercises), and (c) mostly relate to job performance.

Keywords: assessment center; initial impressions; thin slice; rating processes; job performance

Introduction

Every day, people form initial impressions when encountering others. Such impressions are typically based on short slices of verbal and nonverbal behavior of the other person (e.g., Weisbuch et al., 2010), they are formed rapidly (e.g., Willis & Todorov, 2006), and can have far-reaching consequences (see Harris & Garris, 2008; Swider et al., 2022 for an overview). For example, a positive impression can result in evaluating someone as a potential friend, romantic partner, or future employee (e.g., Barrick et al., 2010; Gazzard Kerr et al., 2020; Human et al., 2013; Swider et al., 2016). In personnel selection, initial impressions are thought to be especially relevant for interpersonal selection procedures such as interviews and assessment centers (ACs) and have been found to influence selection outcomes (e.g., Barrick et al., 2010; Ingold et al., 2018). Yet, in contrast to research on interviews scrutinizing the role of initial impressions (Barrick et al., 2010, 2012; Swider et al., 2011, 2016), research on initial impressions in ACs has been comparatively scarce (Swider et al., 2022).

Creating more knowledge on the role of initial impressions in ACs is critical for both practice and theory. Practically, ACs are frequently used across organizations and countries for selecting employees at different levels (Krause & Thornton, 2009), and understanding how impressions

© The Author(s), 2024. Published by Cambridge University Press on behalf of Society for Industrial and Organizational Psychology. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

https://doi.org/10.1017/iop.2024.2 Published online by Cambridge University Press

affect these influential AC ratings is key for designing ACs and training assessors. ACs are highly interpersonal selection procedures during which trained assessors face the demanding task of rating assessees' performance on different constructs in several simulation exercises (International Taskforce on Assessment Center Guidelines, 2015; Kleinmann & Ingold, 2019). Theoretically, this cognitively demanding task creates a condition during which, according to dual process theories (e.g., Evans & Stanovich, 2013), quickly formed impressions should have a strong effect on assessor ratings. Similarly, earlier research has looked at general impressions and their relation to AC ratings (Lance et al., 2004). Empirically, more recent AC research (Ingold et al., 2018) has supported the relevance of initial impressions because it found that initial impressions formed in the first two minutes of an exercise by two different group of raters (i.e., raters with and without assessor experience) reliably predicted AC performance ratings as rated by trained assessors, indicating the need for future research.

The current study aims to contribute to the body of research on the role of initial impressions in ACs. Based on the "thin slices" of behavior paradigm in personality and social psychology (e.g., Ambady & Rosenthal, 1992; Borkenau et al., 2004), we dig into what role the specific slice of the AC exercise (i.e., the time point of the observation that initial impressions are based on) plays for the relationship between initial impressions and selection outcomes. Specifically, we examine effects related to initial impressions when observing a slice of assessees' behavior at either the beginning, middle, *or* end of an AC exercise. Slices from different time points can contain different behavioral samples that assessors can rely on for building their impressions about assessees. On the one hand, assessees might appear prone to conveying constantly a good impression over the course of an AC exercise, making a positive correlation of impressions formed upon different slices likely. On the other hand, in this demanding situation, assessees might vary in their behaviors over the course of an AC exercise due to affective, cognitive, or motivational processes such as nervousness, information processing, or impression management (Fletcher et al., 1997; McFarland et al., 2005).

To provide knowledge on the effects of initial impressions formed from different slices, this study has the following objectives. First, we will explore the consistency of initial impressions formed of assessees across AC exercises (i.e., using slices from the same time points in different exercises) and within AC exercises (i.e., using slices from different time points within the same exercise). This will advance our conceptual understanding of the nature of initial impressions by addressing whether the observed slice of behavior generally matters for forming an initial impression (i.e., does assessees' conveyed impression change across and within exercises). Second, we will scrutinize the relation between initial impressions from different slices and AC performance. This will allow for insights into assessors' cognitive processes by investigating to what extent their ratings are affected by impressions that assessees' convey during exercises, and if these relationships vary. Third, we will explore the criterion-related validity of initial impressions in ACs for different slices. We aim to generate practically relevant knowledge by exploring whether initial impression ratings from different slices contain criterion-relevant or irrelevant variance by examining their correlations with supervisor ratings of assessee's job performance. Apart from increasing the theoretical understanding of impressions conveyed by assessees throughout an exercise, this will inform practitioners on whether to suppress or make use of impressions conveyed from assessees at different time points.

Research questions and hypotheses

Consistency of initial impressions across and within AC exercises

In the evaluative situation of the AC, assessees are supposed to put their best foot forward to achieve favorable ratings, and this should apply across AC exercises as well as to the whole time span of each AC exercise. As such, it appears plausible that the impressions conveyed by assessees

https://doi.org/10.1017/iop.2024.2 Published online by Cambridge University Press

at the same time point (beginning, middle, or end) in different AC exercises, and also at different time points in the same AC exercise, should converge. Findings on initial impressions outside the selection literature support this, indicating their relative stability across different situations (e.g., Borkenau et al., 2004; Leikas et al., 2012).

However, it could be that some assessees behave inconsistently across different exercises depending on the specific situational demands of the exercises (Breil et al., 2023; Lance, 2008; Woehr & Arthur, 2003). This might result in different initial impressions conveyed across different exercises. In a similar vein, it is possible that assessees behave inconsistently within the same exercise. For example, at the beginning of an AC exercise, an assessee might be especially nervous and therefore appear less competent as compared to the end of an AC exercise when the same assessee might be more relaxed and appears more competent. Given the different possibilities, we explore:

Research Question 1. Are initial impressions formed of an assessee based on 2-min observations of the beginning, the middle, and the end of different AC exercises positively related *across* different AC exercises?

Research Question 2. Are initial impressions formed of an assessee based on 2-min observations of the beginning, the middle, and the end of different AC exercises positively related *within* the same AC exercises?

Predicting assessment center performance from initial impressions

From a dual process theory perspective (Evans, 2008; Kahneman, 2003), assessors might be prone to use their quick judgments as a general anchor to arrive at AC performance ratings. Dual process theories assume that two cognitive processes are relevant to human judgment: Type 1 processing that happens automatically (i.e., quick, intuitive, non-conscious judgment) and Type 2 processing that is deliberate (i.e., slower, analytical, conscious judgment; Kahneman, 2003). In an AC, high cognitive demands are placed on assessors, and this represents a fertile ground for intuitive Type 1 processing such as initial impressions to affect judgment outcomes (Kahneman & Frederick, 2002; Stanovich & West, 2000). This is also what has been argued in more recent theorizing on assessor rating processes (Ingold et al., 2018). Empirically, this has only been shown for initial impression based upon thin slices of the beginning of AC exercises (Ingold et al., 2018). Thus, we posit the following hypotheses, and examine Hypothesis 1a as replication:

Hypotheses 1a-c. Initial impressions formed of an assessee based on 2-min observations at the (a) beginning, (b) middle, and (c) end of an AC exercise will predict AC performance ratings as rated by trained assessors in this exercise.

Predicting job performance from initial impressions

A prevailing position in AC textbooks and in AC rater trainings has been hat intuitive Type 1 processing might lead to AC ratings that contain criterion-irrelevant variance (e.g., Thornton et al., 2014), and thus initial impressions might be detrimental to the AC ratings' criterion-related validity and should therefore be avoided. On the other hand, in more recent dual process theory building, a shift has occurred: Whereas past research on dual process theories highlighted the error-proneness of Type's 1 quick and unreflective processes, there is now more attention spent on their advantages (see Evans & Stanovich, 2013; Krueger & Funder, 2004; Kruglanski & Gigerenzer, 2011). Specifically, dual process theory scholars have pointed out that "intuitive thinking can also be powerful and accurate" (Kahneman, 2003, p. 699) and "Type 1 processing can lead to right answers" (Evans & Stanovich, 2013, p.229). For example, decades of research have shown that intuitive judgments of strangers' personality traits are surprisingly accurate

https://doi.org/10.1017/iop.2024.2 Published online by Cambridge University Press

(e.g., Borkenau et al., 2004; Connelly & Ones, 2010). Thus, it seems plausible that initial impressions from ACs contain some criterion-relevant variance and predict job performance. However, the only empirical investigation on this relation points toward a non-significant, positive effect (Ingold et al., 2018), and a recent review on initial impressions in work settings, highlights the ongoing debate about valid versus biased information contained in initial impressions (Swider et al., 2022). Thus, we explore for all three slices:

Research Question 3. How do initial impressions based on 2-min observations of the beginning, the middle, and the end of AC exercises relate to job performance?

Methods

Sample

The original sample consisted of 223 assessees (91 women). This sample was part of a large data collection from the Swiss National Science Foundation (146039; https://data.snf.ch/grants/grant/146039) and has also been studied in Heimann et al. (2022) and Heimann et al. (2021). Assessees were employed in various organizations and had registered to participate in a one-day AC as part of a job application training program to prepare for future applications. They agreed that their data including videotapes of them participating in AC exercises would be used for research purposes. In return for their participation, assessees received feedback and recommendations regarding their AC performance. The mean age of assessees was 30.56 ($SD = 7.32$) years and they had been employed in their current job for 2.57 ($SD = 2.22$) years on average. All assessees provided the contact details of their direct supervisor upon signing up for the job application trainings so that we could collect supervisory job performance ratings. Assessees and assessors did not have access to supervisory performance ratings, and supervisors did not have access to assessees' AC performance ratings.

Procedure

Assessees completed three AC exercises in randomized order: one competitive leaderless group discussion (LGD) and two cooperative LGDs. For each of these interactive exercises, assessees had 15 minutes of individual preparation time and 30 minutes for the actual exercise. In the competitive LGD (Exercise 1), a group of assessees discussed different strategies to solve a given problem. The group had to agree on a rank order for the effectiveness of the different strategies they had discussed, and each assessee had the task to convince the group of their particular rank order. In the two cooperative LGDs (Exercise 2 and Exercise 3), a group of assessees had to find the best solution to a given problem as a team. Each assessee held different pieces of information and they had to share their information with the other assessees in their group to come up with a suitable solution. All three AC exercises had been used in previous AC studies, and had been shown to be criterion-valid (Ingold et al., 2016; Jansen et al., 2013; König et al., 2007). In each AC exercise, a panel of two trained assessors rated assessees' behavioral competencies. Teams of assessors rotated across AC exercises, so that each assessee was evaluated by four assessors (i.e., two teams) in total during the AC. Assessors stemmed from a pool of 43 (34 women) trained university students who were majoring in I-O psychology and who participated as interns to gain experience as assessors. Assessors mean age was 28.33 years ($SD = 6.87$), and they had been studying psychology for 3.20 ($SD = 1.25$) years on average. Two thirds of them (67%) were graduate students. Prior research did not find a difference in the rating accuracy of psychology students versus managers as assessors (Wirz et al., 2013). To comply with recommended AC practices (International Taskforce on Assessment Center Guidelines, 2015), assessors had previously completed a one-day frame-of-reference assessor training (Roch et al., 2012).

https://doi.org/10.1017/iop.2024.2 Published online by Cambridge University Press

After the initial data collection (i.e., the administration of AC exercises), we collected ratings of initial impressions from a novel pool of raters to ensure that assessees' AC performance and initial impressions were evaluated by different individuals. Following the procedure from previous AC research as well as interview research (e.g., Barrick et al., 2010, 2012; Ingold et al., 2018; Schmid Mast et al., 2011; Swider et al., 2011), we obtained initial impression ratings based on 2-min video excerpts. Specifically, we extracted two minutes from the beginning, middle, and end of each of the three video-recorded AC exercises. We had to exclude some videos due to recording issues. In parallel to Ingold et al. (2018), five raters per video independently rated initial impressions. In each AC exercises and at each time point, initial impressions were rated by a different set of five raters—so that each set of raters evaluated each assessee only once. This led to nine different sets of raters (i.e., 45 raters) per assessee. Video clips were randomly assigned to the sets of raters. The rater pool consisted of 150 (114 women) individuals who, similar to the assessors providing AC ratings, were university students majoring in I-O psychology. Their mean age was 24.95 years ($SD = 6.27$), and they had been studying psychology for 1.54 ($SD = 0.96$) years on average. About a quarter of them (26%) were graduate students. We recruited a group of lay raters without requiring any experience as assessors given that prior research showed no difference between ratings from experienced and unexperienced initial impression raters (Ingold et al., 2018). Similar to previous research on initial impressions (Barrick et al., 2010; Swider et al., 2016), raters were not trained and they were not allowed to discuss their ratings with each other as common in the thin slice paradigm. Initial impression raters had no access to any other ratings of assessees (i.e., AC performance, job performance).

Measures

Initial impressions

To assess initial impressions in AC exercises, we used the same scale as Ingold et al. (2018). The scale consisted of eight items targeting raters' general initial impressions of assessees. An example item for this scale is "I have a positive impression of this assessee". Raters rated all items on a 7-point scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). Across raters, AC exercises, and time points the internal consistency of the initial impression scale was $\alpha = .98$. Interrater reliabilities for single raters were ICC(1, 1) = .26 in Exercise 1, .35 in Exercise 2, .30 in Exercise 3, and .30 averaged across exercises. Given that reliabilities for single raters are typically low when assessing initial impressions, each assessee was rated by five raters on each item (see also Swider et al., 2016). Interrater reliabilities of initial impression ratings averaged across the five raters were ICC(1, 5) = .64 in Exercise 1, .73 in Exercise 2, .68 in Exercise 3, and .68 averaged across exercises, which is in line with what to expect based on previous initial impression studies with a similar design (Ingold et al., 2018; Swider et al., 2016).

AC performance

In the AC, assessees' performance was rated on five AC dimensions in total. The AC dimensions were *Communication, Influencing Others, Consideration, Organizing and Planning,* and *Problem Solving* (see Arthur et al., 2003). For each AC exercise, assessors were provided with behavioral anchors for each of the AC dimensions they had to rate. Assessors individually rated four AC dimensions on a 5-point-scale from 1 (*very low expression*) to 5 (*very high expression*) in each of the three AC exercises. As common in AC practice (International Taskforce on Assessment Center Guidelines, 2015), each assessee was rated by a team of two trained assessors in each exercise. After completion of all AC exercises, assessors discussed their individual ratings. To obtain scores of AC performance, we averaged ratings across AC dimensions within each AC exercise. Before discussion, interrater reliabilities of AC ratings from individual assessors were ICC (1, 1) = .64 in Exercise 1, .70 in Exercise 2, .69 in Exercise 3, and .68 averaged across exercises, and interrater reliabilities for AC ratings averaged across the two assessors were ICC(1, 2) = .78 in Exercise 1, .83 in Exercise 2, .81 in

https://doi.org/10.1017/iop.2024.2 Published online by Cambridge University Press

Table 1. Means, Standard Deviations, and Intercorrelations of Initial Impressions in Assessment Center Exercises

Impression	*N*	*M*	*SD*	2	3	4	5	6	7	8	9
1. Exercise 1: Beginning	213	4.66	0.84	**.20**	**.19**	**.19**	**.23**	**.23**	**.26**	**.16**	.07
2. Exercise 1: Middle	213	4.73	0.80		**.27**	**.24**	**.25**	**.21**	**.16**	**.17**	**.16**
3. Exercise 1: End	213	4.63	0.80			**.22**	**.19**	**.15**	**.30**	**.33**	**.17**
4. Exercise 2: Beginning	207	4.54	0.87				**.20**	**.19**	**.28**	**.28**	**.20**
5. Exercise 2: Middle	207	4.67	0.88					**.17**	**.30**	**.22**	**.24**
6. Exercise 2: End	207	4.64	0.82						.14	**.25**	.13
7. Exercise 3: Beginning	197	4.59	0.83							**.28**	**.17**
8. Exercise 3: Middle	197	4.68	0.82								**.34**
9. Exercise 3: End	197	4.75	0.77								

Note. Initial impressions were rated by five initial impression raters. For each assessment center exercise (Exercise 1, 2, and 3) and each time point (beginning, middle, and end), a different set of five raters were used. Significant values ($p < .05$) are displayed in bold.

Exercise 3, and .81 averaged across exercises. After discussion, interrater reliabilities for individual assessors were ICC(1, 1) = .80 in Exercise 1, .82 in Exercise 2, .81 in Exercise 3, and .81 averaged across exercises, and interrater reliabilities for AC ratings averaged across the two assessors were ICC(1, 2) = .89 in Exercise 1, .90 in Exercise 2, .89 in Exercise 3, and .90 averaged across exercises, which is comparable to the interrater reliability reported in previous studies with a similar design (Heimann et al., 2022; Ingold et al., 2016; Jansen et al., 2013; Wirz et al., 2020).

Job performance

After assessees had completed the AC, we contacted their direct supervisors to obtain job performance ratings via an online survey. In total, 201 supervisors (57 women) completed the survey. We measured job performance with nine items from Jansen et al. (2013) on a 7-point scale ranging from 1 (*not at all*) to 7 (*absolutely*). An example item is "S/he achieves the objectives of the job". The internal consistency of the scale was $\alpha = .91$.

Results

All statistical analyses were performed in *R*. We provide the anonymized data and code for our analyses on https://osf.io/e4rsq.

Consistency of initial impressions

Research Question 1 addressed the relationship of initial expressions *across* different AC exercises. Thus, we calculated bivariate correlations of assessees' initial impressions separately for the beginning, middle, and end across the three exercises. Results are displayed in Table 1. Overall, initial impressions from the same time point (beginning, middle, and end of an AC exercise) were consistent *across* AC exercises. For the initial impressions based upon ratings of the beginning of three different exercises, we found significant correlations ranging from $r = .19$–.28 (average correlation .25). Similar results were found for initial impressions based on the middle ($r = .17$–.25; average correlation .21), and the end ($r = .13$–.17; average correlation .15; with one non-significant correlation of .13) of the exercises. Overall, the mean correlation for initial impressions measured at the same time point in different exercises was .20. Hence, individuals who were perceived more favorable on a given slice in one exercise were also perceived more favorable based on the same slices in other exercises.

https://doi.org/10.1017/iop.2024.2 Published online by Cambridge University Press

For Research Question 2, we investigated the relationship of initial impressions from different time points *within* the same AC exercises (see Table 1). All correlations were significant ranging from $r = .19$ (average correlation of initial impressions within Exercise 2) to .26 (average correlation of initial impressions within Exercise 3). The mean correlation for initial impressions measured at different time points within the same exercise was .22. Thus, initial impressions were also consistent within exercises. This may indicate that individuals that were perceived more favorable based on their impression from the beginning of an exercise were also received more favorable based on their impression from the middle and end of the same exercise.

To further explore the consistency of initial impressions, we also investigated the relationships of initial impressions from different exercises *and* different time points (i.e., slices). Correlations ranged from $r = .07$–.33. The mean correlation for initial impressions measured at different time points and in different exercises was .22, and thus similar to the mean correlations for initial impressions measured at the same time point in different exercises, and for initial impressions measured at different time points within the same exercise. Taken together, results speak in favor of some general consistency in initial impressions across different types of slices (exercises and time points), especially given that initial impression ratings of different slices stemmed from different groups of untrained raters.

Initial impressions predicting AC performance

In Hypotheses 1a–c, we predicted that initial impressions formed at the (a) beginning, (b) middle, and (c) end of an AC exercise by untrained raters would predict AC performance ratings as rated by trained assessors in this exercise. To account for the fact that multiple different assessors evaluated different assessees across different exercises, we calculated multilevel models (crossed-random effects; see also Judd et al., 2012 for advantages; Putka & Hoffman, 2013) using the *R* package lme4 (Bates et al., 2014). For this, we specified assessees, assessors, and exercises as random effects. This allows to acknowledge for the fact that (1) some assessees performed better than others ("assessee variance"), (2) some assessors gave more lenient/severe performance ratings than others ("assessor variance"), and (3) some exercises were harder/easier than others ("exercise variance"). As shown in Table 2, while the differences between assessees were large, the differences between assessors or exercises were negligible. Variance in performance ratings that could not be explained by either general assessee performance, assessor rating styles, or exercises effects is found in the residual variance. For our main research question, we calculated three models in which we added initial impressions received from the beginning (see Model 1), the middle (see Model 2), or the end of the AC exercise (see Model 3) as fixed effect (i.e., one model for each time point). Here we wanted to see if higher initial impression ratings (across all assessees, assessors, and exercises) would lead to higher performance ratings. Results showed that initial impressions from all time points significantly predicted AC performance ratings (beginning $b = .17$; $p < .001$; middle $b = .15$; $p < .001$; end $b = .17$; $p < .001$). That is, initial impression ratings were positively related to AC performance, thereby supporting Hypotheses 1a–c.

As part of an exploratory analysis, we further investigated the unique importance of the three initial impressions for AC performance ratings based on the different slices (i.e., from different time points). For this, we added all three initial impressions from different time points as fixed effects into the same model (see Model 4 in Table 2) such that, in contrast to the previous models, all initial impressions were considered simultaneously. All three impressions were significantly related to AC performance, even when controlling for the respective other impressions (beginning $b = .17$; middle $b = .14$; end $b = .16$; all ps $< .001$). Hence, for example, the initial impression received from the end predicted additional AC performance in comparison to the other initial impressions from the other slices (i.e., from the beginning and the middle).

Finally, as additional robustness check, we calculated bivariate correlations for all relevant variables, as can be seen in Table 3. The first six rows in Table 3 refer to initial impressions

https://doi.org/10.1017/iop.2024.2 Published online by Cambridge University Press

Table 2. Multilevel Models of Initial Impressions Predicting Assessment Center Ratings

Variables	Model 1	Model 2	Model 3	Model 4
Random effect: Assessee variance	.29	.28	.29	.21
Random effect: Assessor variance	.02	.02	.02	.02
Random effect: Exercise variance	.01	.01	.01	.01
Residual variance	.25	.25	.25	.24
Fixed effect: Initial impressions beginning	.17 ($p < .001$)			.17 ($p < .001$)
Fixed effect: Initial impressions middle		.15 ($p < .001$)		.14 ($p < .001$)
Fixed effect: Initial impressions end			.17 ($p < .001$)	.16 ($p < .001$)

Note. Displayed are four different multilevel models with three random intercepts each (as well as the residual variance). Assessee variance: Variance in performance ratings attributable to different assessees (i.e., some assessees performed better than others). Assessor variance: Variance in performance ratings attributable to different assessors (i.e., some assessors gave more lenient/severe performance ratings than others). Exercise variance: Variance in performance rating attributable to different exercises (i.e., some exercises were harder/easier than others). Residual variance: Variance in performance ratings that could not be explained by assessee, assessor, or exercise main effects. The included fixed effects differed depending on the model: Model 1 includes only initial impressions based on the beginning of the assessment center (AC) exercises, Model 2 includes only initial impressions based on the middle of the AC exercises, Model 3 includes only initial impressions based on the end of the AC exercises, and Model 4 includes initial impressions from all time points.

aggregated to the time point (Rows 1 to 3) or the exercise (Rows 4–6). Row 7 refers to an overall initial impression rating aggregated across time points and exercises. Thus, correlations between the first six rows and Row 7 can be seen as part-whole correlation. Rows 8–11 refer to the AC performance ratings (on the exercise-level and the overall-level). Row 12 refers to job performance. The main results presented in this table (i.e., correlations between initial impressions from the beginning, middle, and end of an AC exercise with performance in the respective exercise) are in line with the results of the crossed-random effects models that are presented in Table 2.

Initial impressions predicting job performance

Research Question 3 explored the relationship between initial impressions and supervisor-rated job performance (see Table 3 for results). We performed this analysis with initial impressions aggregated (1) across exercises, (2) across time points, and (3) across exercises and time points. First, we analyzed initial impressions aggregated across the three exercises for each time point. Results of bivariate correlations showed that initial impressions based on the middle and on the end of AC exercises significantly related to job performance (middle $r = .16$, $p = .030$; end $r = .17$, $p = .019$). Initial impressions based on the beginning of the exercises did not significantly relate to job performance ($r = .11$; $p = .146$). Correlations between initial impressions from different time points and job performance did not differ significantly from each other, neither for initial impression from the beginning versus the middle ($z = -0.68$, $p = .252$), the beginning versus the end ($z = -0.78$, $p = .217$), nor the middle versus the end ($z = -0.15$, $p = .443$).

Second, we aggregated initial impressions across the three time points for each exercise. Here, initial impressions correlated significantly with job performance for Exercise 1 ($r = .22$, $p = .002$) and Exercise 3 ($r = .21$, $p = .006$), but not for Exercise 2 ($r = .07$, $p = .377$), even though all three exercises were criterion-valid (see Table 3). Correlations between initial impressions from different exercises and job performance differed significantly from each other for Exercise 1 versus Exercise 2 ($z = 2.04$, $p = .020$), and for Exercise 2 versus Exercise 3 ($z = -1.81$, $p = .035$), but not for Exercise 1 versus Exercise 3 ($z = 0.14$, $p = .445$).

https://doi.org/10.1017/iop.2024.2 Published online by Cambridge University Press

Table 3. Means, Standard Deviations, and Intercorrelations of Main Study Variables

Variable	*N*	*M*	*SD*	2	3	4	5	6	7	8	9	10	11	12
1. Initial: Beginning	213	4.60	0.61	**.43**	**.40**	**.63**	**.65**	**.60**	**.78**	**.33**	**.41**	**.46**	**.48**	.11
2. Initial: Middle	213	4.70	0.58		**.51**	**.64**	**.67**	**.65**	**.82**	**.48**	**.50**	**.51**	**.59**	**.16**
3. Initial: End	213	4.67	0.53			**.60**	**.62**	**.65**	**.78**	**.44**	**.46**	**.48**	**.55**	**.17**
4. Initial: Exercise 1	213	4.67	0.56				**.45**	**.40**	**.78**	**.52**	**.40**	**.42**	**.54**	**.22**
5. Initial: Exercise 2	207	4.61	0.58					**.47**	**.82**	**.35**	**.55**	**.45**	**.54**	.07
6. Initial: Exercise 3	197	4.67	0.57						**.79**	**.36**	**.43**	**.58**	**.55**	**.21**
7. Initial: Overall	213	4.65	0.45							**.52**	**.58**	**.60**	**.68**	**.18**
8. AC: Exercise 1	223	3.51	0.72								**.42**	**.46**	**.75**	**.23**
9. AC: Exercise 2	223	3.64	0.75									**.74**	**.87**	**.21**
10. AC: Exercise 3	223	3.50	0.76										**.88**	**.15**
11. AC: Overall	223	3.55	0.62											**.23**
12. Job performance	201	5.99	0.82											

Note. Initial impressions were rated by initial impression raters. Initial: beginning = Initial impression ratings from the beginning aggregated across exercises. Initial: middle = Initial impression ratings from the middle aggregated across exercises. Initial: end = Initial impression ratings from the end aggregated across exercises. Initial: Exercise 1 = Initial impression ratings from Exercise 1 aggregated across time points. Initial: Exercise 2 = Initial impression ratings from Exercise 2 aggregated across time points. Initial: Exercise 3 = Initial impression ratings from Exercise 3 aggregated across time points. Initial: Overall = Initial impression ratings aggregated across exercises and time points. Assessment center (AC) ratings were rated by two assessors. AC: Exercise 1 = Assessment center ratings aggregated across assessors at Exercise 1. AC: Exercise 2 = Assessment center ratings aggregated across assessors at Exercise 2. AC: Exercise 3 = Assessment center ratings aggregated across assessors at Exercise 3. AC: Overall = Assessment center ratings aggregated across assessors and exercises. Job performance = Job performance ratings by supervisors. Significant values ($p < .05$) are displayed in bold.

Lastly, when aggregating initial impressions across all exercises and time points, the relationship between overall initial impressions and job performance was significant ($r = .18$, $p = .012$). Descriptively, this relationship appeared to approach the relationship between aggregated ratings of AC performance and job performance ($r = .23$, $p < .001$). Statistically, the correlations of job performance with overall initial impressions and with overall AC performance did not differ significantly from each other ($z = -0.90$, $p = .184$).

Discussion

The present study set out to extend prior research highlighting the role of initial impressions for AC ratings (Ingold et al., 2018) and to generate new knowledge about the consistency and relevance of slice positions for impressions in ACs. Drawing from the "thin slices" of behavior paradigm (e.g., Ambady, 2010; Ambady & Rosenthal, 1992), results showed that initial impressions formed upon different slices of AC exercises (i.e., two minutes at the beginning, middle, and end of an AC exercise) were consistent across and within exercises, predicted AC performance, and—for two out of three time points and exercises—also predicted job performance.

Findings and implications for theory and research

A key finding of this study is that the relationship between initial impressions and AC performance is consistent and independent of what slice of an AC exercise is used to assess initial impressions. Across different exercises, we found that initial impressions were significantly related to ratings of AC performance—no matter if initial impressions were based on observing assessees at the beginning, middle, or end of an exercise. This also aligns with our finding that initial impressions were relatively consistent across different slices (i.e., across exercises and time points).

https://doi.org/10.1017/iop.2024.2 Published online by Cambridge University Press

Thus, assessees seem to convey similar impressions across different slices that constantly relate to AC performance. This effect exists even across different raters (as every rater was only allowed to see a candidate once) which speaks in favor of assessees' consistency in conveying impressions. Conceptually, these findings imply that the initial impressions that a person evokes in others are (1) a relatively stable person characteristics in different evaluative situations within an AC (consistent with research on social cognition; e.g., Fiske et al., 2007), and (b) that, across different slices, these impressions are likely to influence how assessors evaluate this person's behavior in an AC (consistent with dual process theory; Evans, 2008).

A second relevant finding is that most of the initial impression ratings were positively related to job performance, thereby suggesting that initial impressions appear to contain some criterion-relevant information. Concerning the relevance of the time point of the slice in relation to job performance, the timing of the slice appeared irrelevant for the prediction of job performance, given that the consistently positive correlations from different time points did not differ statistically. At the same time, the relationship of exercise-level initial impressions and job performance appeared to differ across exercises (with one correlation based on initial impressions in one exercise being not significant and differing in size in comparison to the two correlations from the other two exercises). Thus, it seems noteworthy to study the influence of the situation on the criterion-relevance of initial impressions systematically (see below)—especially since all three exercises were structurally similar (i.e., all of them were LGDs) and criterion-valid (i.e., ratings of AC performance from all three exercises related to supervisory performance ratings). In combination with the exploratory finding that the relationship between overall initial impressions (i.e., when aggregated across all time points and exercises) and job performance did not differ significantly from the relationship of AC performance with job performance, these findings provide further evidence for the perspective that initial impressions in interactive personnel selection methods can contain valid information (Barrick et al., 2010; Ingold et al., 2018).

At first sight, our results might then be read to suggest that when raters can observe assessees' behavior in well-constructed AC exercises, detailed observations by trained assessors could be accompanied and to some degree potentially compensated for by using a *larger* amount of untrained raters providing generic evaluations (i.e., initial impression ratings). Yet, the reliability and total number of raters for AC ratings versus initial impression raters needs to be taken into consideration when interpreting the criterion-related validity results of aggregated initial impression ratings. The reliability of a single untrained lay rater in this study was ICC (1,1) = .30, whereas the reliability of a single trained assessor was ICC (1,1) = .68 (before discussing their individual ratings with other assessors). This illustrates that ratings from untrained lay raters are likely to be substantially less reliable than ratings from trained assessors, which in turn implies that a lot more lay raters are needed to achieve reliable scores (see also Wirz et al., 2013). For example, the overall score for AC performance in the present study relied on ratings from four assessors per assessee, whereas the overall score for initial impressions relied on ratings from 45 raters per assessee. In line with classical test theory (e.g., Gulliksen, 1950), increasing the number of raters (i.e., which is equivalent to increasing test length) enhances the reliability of a measure, which has the potential to improve validity (if performance-relevant information is assessed) such that the considerable higher number of raters worked in favor of the criterion-related validity of initial impressions. This favorable effect of aggregating ratings from multiple raters has also been discussed in the context of initial impressions (Eisenkraft, 2013; Ingold et al., 2018) as well as for multiple shorter speed assessments (Herde & Lievens, 2023). Accordingly, we see this research as a starting point and advocate for more research that scrutinizes the factors contributing to the criterion-relevant variance of initial impression ratings to allow for a more comprehensive database and a better theoretical understanding of the underlying factors.

To advance from here, future research particularly needs to shed light on the key drivers of the criterion-relevance of initial impressions, thereby extending prior research on antecedents of initial impressions in selection settings (Barrick et al., 2012; Ingold et al., 2018; Swider et al., 2016).

https://doi.org/10.1017/iop.2024.2 Published online by Cambridge University Press

One methodological approach toward addressing this lies in coding micro-behaviors (Breil et al., 2023) that are potentially observable in thin slices of AC exercises. This would allow answering the question to what extent initial impressions rely on observations of micro-behaviors (i.e., the degree to which they may be influenced by Type 2 processing; intentional and targeted evaluations), and thereby contribute to assessors' ratings of AC performance, respectively. This is also relevant in light of the discussion about how Type 1 and Type 2 processing interrelate (Evans & Stanovich, 2013), and in light of determining why some initial impressions appear to be more job relevant than others (as discussed above).

Whereas selection research on initial impression has tended to look into the characteristics of the person being rated (e.g., Barrick et al., 2012; Ingold et al., 2018), another relevant area of research relates to shedding light on the role of situational characteristics on the criterion-relevance of initial impressions. In this study, the interpersonal, work-related nature of the AC exercises might have contributed to the criterion-relevance of initial impressions. To allow for a better understanding of the situational component, we suggest shedding light on how the fidelity and job relevance impacts the degree to which initial impressions may predict job performance, for instance by manipulating situational characteristics in experimental studies.

Limitations

A relevant limitation is that the AC in this study was not conducted for selection purposes, which could overestimate the criterion-related validity of initial impressions to some degree. The AC in this study was part of a job application training and therefore stakes were not as high as for assessees in real selection situations—even though we simulated an AC that followed the international AC guidelines (International Taskforce on Assessment Center Guidelines, 2015), and assessees reported that they were motivated to perform well in the AC. In a high-stakes selection AC, assessees might have been more nervous or would engage in more impression management as compared to the present AC. In consequence, the initial impressions that assessees convey in a high-stakes AC might be less criterion-valid than the initial impressions conveyed in the present AC because assessees' appearance and behavior in a high-stakes AC might be less representative of the impressions they typically convey at work.

A second limitation is that assessees in the present study held various jobs, which could underestimate the criterion-related validity of AC ratings to some degree. Although the conducted AC was criterion-valid, criterion-validity estimates for AC performance ratings might potentially be higher in a study with assessees who hold similar types of jobs and in which AC exercises are specifically tailored to this type of job (see Bartels & Doverspike, 1997).

Implications for AC practices

The findings of this study offer two major takeaways for AC practitioners. First, initial impressions appear to contain information that is relevant to job performance. Therefore, it might be helpful to consider assessing initial impressions in ACs (notably in addition to traditional AC ratings)—instead of telling raters to suppress initial impressions. Second, in light of earlier research on the relevance of different exercises for the criterion-related validity (Speer et al., 2014), practitioners and researchers should join forces in piloting ACs with multiple shortened exercises (Herde & Lievens, 2020), and notably evaluate effects on the AC's criterion-related validity across various samples to allow for meta-analytic conclusions on the validity of multiple shorter observations with different rating approaches in ACs in the long run.

Data availability statement. The data and R code for this article can be found in the online supplements at https://osf.io/e4rsq.

https://doi.org/10.1017/iop.2024.2 Published online by Cambridge University Press

Acknowledgements. We thank Nadine Maier und Anna Luisa Grimm for helping to collect the initial impression data for this study. Parts of the data were collected as part of a larger Swiss National Science Foundation Project (grant number 146039), and parts of the these data have been used in Heimann et al. (2022) and Heimann et al. (2021).

Competing interests. We have no known conflict of interest to disclose.

References

Ambady, N. (2010). The perils of pondering: Intuition and thin slice judgments. *Psychological Inquiry*, **21**(4), 271–278. Doi: 10.1080/1047840X.2010.524882.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, **111**(2), 256–274. Doi: 10.1037/0033-2909.111.2.256.

Arthur, W., Day, E. A., McNelly, T. L., & Edens, P. S. (2003, A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology*, **56**(1), 125–154. Doi: 10.1111/j.1744-6570.2003.tb00146.x.

Barrick, M. R., Dustin, S., Giluk, T., Stewart, G., Shaffer, J., & Swider, B. (2012). Candidate characteristics driving initial impressions during rapport building: Implications for employment interview validity. *Journal of Occupational and Organizational Psychology*, **85**(2), 330–352. Doi: 10.1111/j.2044-8325.2011.02036.x.

Barrick, M. R., Swider, B. W., & Stewart, G. L. (2010). Initial evaluations in the interview: Relationships with subsequent interviewer evaluations and employment offers. *Journal of Applied Psychology*, **95**(6), 1163–1172. Doi: 10.1037/A0019918.

Bartels, L. K., & Doverspike, D. (1997). Effects of disaggregation on managerial assessment center validity. *Journal of Business and Psychology*, **12**, 45–53. Doi: 10.1023/A:1025010116265.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**(1), 1–48. Doi: 10.18637/jss.v067.i01.

Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology*, **86**(4), 599–614. Doi: 10.1037/0022-3514.86.4.599.

Breil, S. M., Lievens, F., Forthmann, B., & Back, M. D. (2023). Interpersonal behavior in assessment center role-play exercises: Investigating structure, consistency, and effectiveness. *Personnel Psychology*, **76**(3), 759–795. Doi: 10.1111/peps.12507.

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, **136**(6), 1092–1122. Doi: 10.1037/a0021212.

Eisenkraft, N. (2013). Accurate by way of aggregation: Should you trust your intuition-based first impressions? *Journal of Experimental Social Psychology*, **49**(2), 277–279. Doi: 10.1016/j.jesp.2012.11.005.

Evans, J. S. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, **59**, 255–278. Doi: 10.1146/annurev.psych.59.103006.093629.

Evans, J. S., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, **8**(3), 223–241. Doi: 10.1177/1745691612460685.

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, **11**(2), 77–83. Doi: 10.1016/j.tics.2006.11.005.

Fletcher, C., Lovatt, C., & Baldry, C. (1997). A study of state, trait, and test anxiety, and their relationship to assessment center performance. *Journal of Social Behavior & Personality*, **12**(5), 205–215.

Gazzard Kerr, L., Tissera, H., McClure, M. J., Lydon, J. E., Back, M. D., & Human, L. J. (2020). Blind at first sight: The role of distinctively accurate and positive first impressions in romantic interest. *Psychological Science*, **31**(6), 715–728. Doi: 10.1177/0956797620919674.

Gulliksen, H. (1950). Effect of test length on reliability (general case). In H. Gulliksen (Eds.), *Theory of mental tests* (pp. 74–87). John Wiley & Sons Inc.

Harris, M. J., & Garris, C. P. (2008). You never get a second chance to make a first impression: Behavioral consequences of first impressions. In N. Ambady, & J. J. Skowronski (Eds.), *First impressions* (pp. 147–168). Guilford Publications.

Heimann, A. L., Ingold, P. V., Debus, M. E., & Kleinmann, M. (2021). Who will go the extra mile? Selecting organizational citizens with a personality-based structured job interview. *Journal of Business and Psychology*, **36**(4), 985–1007. Doi: 10.1007/s10869-020-09716-1.

Heimann, A. L., Ingold, P. V., Lievens, F., Melchers, K. G., Keen, G., & Kleinmann, M. (2022). Actions define a character: Assessment centers as behavior-focused personality measures. *Personnel Psychology*, **75**(3), 675–705. Doi: 10.1111/peps.12478.

Herde, C. N., & Lievens, F. (2023). Multiple, speeded assessments under scrutiny: Underlying theory, design considerations, reliability, and validity. *Journal of Applied Psychology*, **8**(3), 351–373. Doi: 10.1037/apl0000603.

Herde, C. N., & Lievens, F. (2020). Multiple speed assessments: Theory, practice, and research evidence. *European Journal of Psychological Assessment*, **36**(2), 237–249. Doi: 10.1027/1015-5759/a000512.

https://doi.org/10.1017/iop.2024.2 Published online by Cambridge University Press

Human, L. J., Sandstrom, G. M., Biesanz, J. C., & Dunn, E. W. (2013). Accurate first impressions leave a lasting impression: The long-term effects of distinctive self-other agreement on relationship development. *Social Psychological and Personality Science*, **4**(4), 395–402. Doi: 10.1177/1948550612463735.

Ingold, P. V., Dönni, M., & Lievens, F. (2018). A dual-process theory perspective to be`tter understand judgments in assessment centers: The role of initial impressions for dimension ratings and validity. *Journal of Applied Psychology*, **103**(12), 1367–1378. Doi: 10.1037/apl0000333.

Ingold, P. V., Kleinmann, M., König, C. J., & Melchers, K. G. (2016). Transparency of assessment centers: Lower criterion-related validity but greater opportunity to perform? *Personnel Psychology*, **69**(2), 467–497. Doi: 10.1111/peps.12105.

International Taskforce on Assessment Center Guidelines (2015). Guidelines and ethical considerations for assessment center operations. *Journal of Management*, **41**(4), 1244–1273. Doi: 10.1177/0149206314567780.

Jansen, A., Melchers, K. G., Lievens, F., Kleinmann, M., Brändli, M., Fraefel, L., & König, C. J. (2013). Situation assessment as an ignored factor in the behavioral consistency paradigm underlying the validity of personnel selection procedures. *Journal of Applied Psychology*, **98**(2), 326–341. Doi: 10.1037/a0031257.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, **103**(1), 54–69. Doi: 10.1037/a0028347.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, **58**(9), 697–720. Doi: 10.1037/0003-066X.58.9.697.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). Cambridge University Press.

Kleinmann, M., & Ingold, P. V. (2019). Toward a better understanding of assessment centers: A conceptual review. *Annual Review of Organizational Psychology and Organizational Behavior*, **6**(1), 349–372. Doi: 10.1146/annurev-orgpsych-012218-014955.

König, C. J., Melchers, K. G., Kleinmann, M., Richter, G. M., & Klehe, U.-C. (2007). Candidates' ability to identify criteria in nontransparent selection procedures: Evidence from an assessment center and a structured interview. *International Journal of Selection and Assessment*, **15**(3), 283–292. Doi: 10.1111/j.1468-2389.2007.00388.x.

Krause, D. E., & Thornton, 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**(4), 557–585. Doi: 10.1111/j.1464-0597.2008.00371.x.

Krueger, J. I., & Funder, D. C. (2004). Towards a balanced social psychology: Causes, consequences, and cures for the problem-seeking approach to social behavior and cognition. *The Behavioral and Brain Sciences*, **27**(3), 313–327.

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. *Psychological Review*, **118**(1), 97–109. Doi: 10.1037/a0020762.

Lance, C. E. (2008). Why assessment centers do not work the way they are supposed to. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, **1**(1), 84–97. Doi: 10.1111/j.1754-9434.2007.00017.x.

Lance, C. E., Foster, M. R., Gentry, W. A., & Thoresen, J. D. (2004). Assessor cognitive processes in an operational assessment center. *Journal of Applied Psychology*, **89**(1), 22–35. Doi: 10.1037/0021-9010.89.1.22.

Leikas, S., Lönnqvist, J.-E., & Verkasalo, M. (2012). Persons, situations, and behaviors: Consistency and variability of different behaviors in four interpersonal situations. *Journal of Personality and Social Psychology*, **103**(6), 1007–1022. Doi: 10.1037/a0030385.

McFarland, L. A., Yun, G., Harold, C. M., Viera, L., & Moore, L. G. (2005). An examination of impression management use and effectiveness across assessment center excercises: The role of competency demands. *Personnel Psychology*, **58**(4), 949–980. Doi: 10.1111/j.1744-6570.2005.00374.x.

Putka, D. J., & Hoffman, B. J. (2013). Clarifying the contribution of assessee-, dimension-, exercise-, and assessor-related effects to reliable and unreliable variance in assessment center ratings. *Journal of Applied Psychology*, **98**(1), 114–133. Doi: 10.1037/A0030887.

Roch, S. G., Woehr, D. J., Mishra, V., & Kieszczynska, U. (2012). Rater training revisited: An updated meta-analytic review of frame-of-reference training. *Journal of Occupational and Organizational Psychology*, **85**(2), 370–395. Doi: 10.1111/j.2044-8325.2011.02045.x.

Schmid Mast, M., Bangerter, A., Bulliard, C., & Aerni, G. (2011). How accurate are recruiters' first impressions of applicants in employment interviews? *International Journal of Selection and Assessment*, **19**(2), 198–208. Doi: 10.1111/j.1468-2389.2011.00547.x.

Speer, A. B., Christiansen, N. D., Goffin, R. D., & Goff, M. (2014). Situational bandwidth and the criterion-related validity of assessment center ratings: Is cross-exercise convergence always desirable? *Journal of Applied Psychology*, **99**(2), 282–295. Doi: 10.1037/a0035213.

Stanovich, K. E., & West, R. F. (2000). Advancing the rationality debate. *Behavioral and Brain Sciences*, **23**(5), 701–717. Doi: 10.1017/S0140525X00623439.

https://doi.org/10.1017/iop.2024.2 Published online by Cambridge University Press

Swider, B. W., Barrick, M. R., & Harris, T. B. (2016). Initial impressions: What they are, what they are not, and how they influence structured interview outcomes. *Journal of Applied Psychology*, **101**(5), 625–638. Doi: 10.1037/apl0000077.

Swider, B. W., Barrick, M. R., Harris, T. B., & Stoverink, A. C. (2011). Managing and creating an image in the interview: The role of interviewee initial impressions. *Journal of Applied Psychology*, **96**(6), 1275–1288. Doi: 10.1037/a0024005.

Swider, B. W., Harris, T. B., & Gong, Q. (2022). First impression effects in organizational psychology. *Journal of Applied Psychology*, **107**(3), 346–369. Doi: 10.1037/apl0000921.

Thornton, G. C., Rupp, D. E., & Hoffman, B. J. (2014). *Assessment center perspectives for talent management strategies*.

Weisbuch, M., Ambady, N., Clarke, A. L., Achor, S., & Weele, J. V.-V. (2010). On being consistent: The role of verbal-nonverbal consistency in first impressions. *Basic and Applied Social Psychology*, **32**(3), 261–268. Doi: 10.1080/01973533.2010.495659.

Willis, J., & Todorov, A. (2006). First impressions making up your mind after a 100-ms exposure to a face. *Psychological Science*, **17**(7), 592–598. Doi: 10.1111/j.1467-9280.2006.01750.x.

Wirz, A., Melchers, K. G., Kleinmann, M., Lievens, F., Annen, H., Blum, U., & Ingold, P. V. (2020). Do overall dimension ratings from assessment centres show external construct-related validity? *European Journal of Work and Organizational Psychology*, **29**(3), 405–420. Doi: 10.1080/1359432X.2020.1714593.

Wirz, A., Melchers, K. G., Lievens, F., De Corte, W., & Kleinmann, M. (2013). Trade-offs between assessor team size and assessor expertise in affecting rating accuracy in assessment centers. *Journal of Work and Organizational Psychology*, **29**(1), 13–20. Doi: 10.5093/tr2013a3.

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**(2), 231–258. Doi: 10.1177/014920630302900206.

Cite this article: Ingold, P. V., Heimann, A. L., & Breil, S. M. (2024). Any slice is predictive? On the consistency of impressions from the beginning, middle, and end of assessment center exercises and their relation to performance. *Industrial and Organizational Psychology* **17**, 192–205. https://doi.org/10.1017/iop.2024.2

https://doi.org/10.1017/iop.2024.2 Published online by Cambridge University Press