

FOCAL ARTICLE

Perceptions of assessment center exercises: Between exercises differences and interventions

Sylvia G. Roch

University at Albany, State University of New York, College of Arts and Sciences, Albany, NY, USA
Email: sroch@albany.edu

(Received 21 December 2023; accepted 3 January 2024; first published online 18 March 2024)

Abstract

Preliminary research has demonstrated that not all assessment center (AC) exercises are viewed as equally just or motivating. The current research builds upon this research and investigates the relationships between six AC exercises and perceptions of self-efficacy, motivation, assessor bias, and fairness. Using a $2 \times 2 \times 2$ experimental design (two informational justice interventions and one rating timing intervention), 286 working adults completed a survey designed to investigate differences between AC exercises and to investigate interventions designed to influence AC exercise perceptions. The results show not only significant perceptual differences between assessor-rated exercises and an ability test but also differences among the rated exercises. The results suggest that an ability test can be perceived as both among the most just and motivating exercises. Lastly, even though the experimental interventions did not have their anticipated effects, the results suggest benefits to having assessors rate recorded participant behaviors versus rating “live” behaviors, benefits that to a certain extent depend on whether participants had previously attended an assessment center.

Keywords: Assessment centers; applicant reactions; applicant motivation

Introduction

Assessment centers are widely used to assess thousands of people each year, mostly for recruitment, selection, and development purposes (e.g., Krause & Thornton, 2009; Lanik, 2019). Surprisingly, few researchers have focused on reactions to specific assessment center (AC) exercises (see Roch et al., 2008; 2014 for exceptions). Researchers have investigated reactions to various selection methods (see the Hausknecht et al., 2004, meta-analysis), but ACs differ from other selection methods in that they consist of several exercises representing several selection methods, some of which have been investigated in the context of reactions to selection methods (e.g., work sample) and some not (e.g., leaderless group discussion).

Understanding how participants' reactions can fluctuate from exercise to exercise within an AC has benefits, both in terms of AC design and in increasing our understanding of why participants may not perform consistently across exercises. Our preliminary research (Roch et al., 2008; 2014) demonstrated that participants' justice perceptions and motivation levels differ among AC exercises. Justice perceptions have been linked to many organizational attitudes and behaviors (Colquitt et al., 2013) and acceptance of job offers (Harold et al., 2016). Motivation has been linked to performance regarding both cognitive ability tests (e.g., Chan et al., 1997; Ployhart & Ehrhart, 2002) and interviews (e.g., Maurer et al., 2001).

Perhaps by understanding why motivation and justice perceptions fluctuate between AC exercises, practitioners can design interventions to bring the less motivating AC exercises in line

with the more motivating ones, making it easier for assessors to identify participants' strengths and weaknesses. Thus, one purpose is to replicate and expand upon our preliminary research (Roch et al., 2008; 2014) investigating differing reactions to AC exercises.

The second purpose is to investigate three interventions, two interventions related to informational justice, focusing on rating explanations (Truxillo et al., 2009), and a third, more applied intervention, focusing on whether exercises are rated immediately or recorded and rated later. The Truxillo et al. (2009) meta-analysis investigating explanations in a selection context found that providing explanations was positively related to both fairness perceptions and test taking motivation. The rating timing intervention is based on a change to the AC standards. The 2015 standards include the practice of recording AC exercises and rating them later (Rupp et al., 2015), which represents a change from the 2000 guidelines, which stressed that AC exercises should be observed by assessors as they occur (Joiner, 2000).

Perceptions of AC exercises

Selection researchers have found that not all selection methods are perceived similarly (e.g., Hausknecht et al., 2004; Hülshager & Anderson, 2009; Kravitz et al., 1996; Smither et al., 1993; Steiner & Gilliland, 1996). Steiner and Gilliland (1996) asked participants to view and rate the procedural justice and perceived favorability of 10 selection methods. Other researchers around the world have adopted Steiner and Gilliland's approach (e.g., Anderson & Witvliet, 2008; Phillips & Gully, 2002; Moscoso & Salgado, 2004), finding only minor differences among countries (Anderson & Witvliet, 2008). Some of these selection methods are commonly used as AC exercises, such as work sample tests, and others are less commonly used but can be found in ACs, such as cognitive ability tests and interviews (Thornton & Rupp, 2006).

Interestingly, Snyder and Shahani-Denning (2012) used the Steiner and Gilliland (1996) procedure and included ACs. They found that ACs received the 6th highest favorability rating out of the 12 rated selection methods. Researchers have also investigated participant reactions to the ACs that they attended and generally found that ACs are perceived positively (e.g., König et al., 2015; Macan et al., 1994; Merkulova et al., 2014; Rupp et al., 2006). However, not all AC exercises may be viewed equally positively. In previous studies (Roch et al., 2008; 2014), we investigated perceptions of a cognitive ability test, semi-structured interview, and written role play in AC simulations using undergraduate students. We proposed that, according to Leventhal's (1980) criteria for procedural justice, cognitive ability tests should be seen as more procedurally just than rated exercises because they are scored based on factually correct answers, which we found indeed to be the case.

In our previous research (Roch et al., 2014), we noted that Steiner and Gilliland (1996) and others using their procedural justice questionnaire found that interviews were viewed as more procedurally just than written ability tests. We proposed that these contradictory results can be attributed to different operationalizations of procedural justice. We used Colquitt's procedural justice measure focusing on Leventhal's (1980) procedural justice criteria. Steiner and Gilliland (1996) used a conglomerate procedural justice measure focusing on many perceptions, ranging from whether the selection method is based on solid scientific research to whether it is impersonal and cold. Given our adoption of Colquitt's (2001) definition of procedural justice, it is not surprising that our participants viewed cognitive ability tests as more procedurally just than the rated exercises. Surprisingly, the rated exercises were viewed as *more* motivating than the cognitive ability test. Procedural justice perceptions have been shown to be positively related to test taking motivation (Bell et al., 2006).

In our previous work (Roch et al., 2014), we offered an explanation using expectancy theory. We proposed that participants may have a higher expectancy that if they put forth effort, they can perform well on rated AC exercises versus more objectively scored ones, hence the higher

motivation levels for rated AC exercises. However, we did not assess expectancy but perceived performance and perceived influence. Perceived influence had not previously caught researchers' attention. However, both Sanchez et al. (2000) and Bell et al. (2006) investigated applicant perceived performance and found a relationship between perceived performance and applicant motivation. We found that both perceived performance and influence were significantly higher for the rated exercises than the cognitive ability test. However, it can be argued that self-efficacy is conceptually closer to expectancy theory than perceived performance. According to Locke et al. (1986), effort-performance expectancy and self-efficacy are closely related. If we are correct and individuals have a higher expectancy that if they put forth effort, they can perform well on rated exercises versus an ability test, this should be reflected in their self-efficacy perceptions.

Thus, the first goal is to replicate our (Roch et al., 2014) main findings regarding motivation and justice but focusing on a general fairness perception instead of procedural justice and also to investigate perceived bias, along with self-efficacy. If an ability test is viewed as more procedurally just than rated exercises, it should also be viewed as less biased, given that perceived bias is represented in Colquitt's (2001) procedural justice measure. Ambrose and Schminke (2009) showed that the justice dimensions (procedural, distributive, interpersonal, and informational justice) relate to organizational outcomes via a general fairness perception.

Hypothesis 1. Individuals will perceive an ability test, containing factually correct answers, as fairer (*Hypothesis 1a*) and less biased (*Hypothesis 1b*) than rated exercises but will report lower exercise self-efficacy (*Hypothesis 1c*) and motivation (*Hypothesis 1d*).

However, not all rated exercises may engender the same reactions, given that exercises can differ in terms of modality (e.g., written case analysis versus oral presentation) and whether participants take part in an exercise individually or as a group (e.g., oral presentation versus leaderless group discussion). This is only speculation and differences between rated exercises will be explored but no hypotheses proposed.

Research Question 1. Are all rated exercises perceived similarly in terms of motivation, fairness, bias, and self-efficacy?

Interventions

Perhaps one way of reducing the perceptual differences between AC exercises is by increasing informational justice perceptions via the use of explanations. In their meta-analysis Truxillo et al. (2009) found that explanations significantly related to fairness perceptions and test motivation. They also investigated moderators. Studies based on scenarios/simulations (19 studies) revealed smaller effect sizes than studies based on actual selection processes (5 studies), but the effects sizes did not include zero in the confidence interval. Furthermore, only a few studies in their meta-analysis focused on a specific selection measure, either a cognitive ability test (six studies) or a personality inventory (six studies). Typical AC exercises were not included in this meta-analysis.

Receiving rating explanations may be especially important in rated AC exercises because there is more ambiguity regarding how assessors rate these exercises versus exercises with demonstrable correct answers. An explanation of the rigorous rating protocol commonly used to rate AC exercises, along with explanations regarding the relevant performance dimensions, may increase informational justice perceptions, given that informational justice focuses on the adequacy of explanations (Colquitt, 2001). Given that informational justice relates to an overall justice perception (Ambrose & Schminke, 2009), interventions based on informational justice may increase AC exercise fairness perceptions and decrease bias perceptions, along with increasing motivation. Two such interventions designed to increase informational justice perceptions by providing explanations will be investigated. Justice perceptions have been positively related to motivation in other contexts (Zapata-Phelan et al., 2009).

Hypothesis 2. Providing explanations regarding the AC rating procedures and the rating dimensions will decrease perceptions of bias (Hypothesis 2a) and increase both fairness perceptions (Hypothesis 2b) and motivation (Hypothesis 2c) across AC exercises.

Truxillo et al. (2009) also investigated self-perceptions, operationalized as a combination of self-efficacy, perceived performance, and self-esteem, but did not find a significant relationship between explanations and self-perceptions. However, they focused on a wide range of explanations. Perhaps if explanations more closely adhered to Colquitt's (2001) definition of informational justice, they may increase self-efficacy. Colquitt et al. (2013) found that informational justice is significantly related to positive state affect. Affect has been related to employee self-efficacy (e.g., Rego et al., 2012).

Hypothesis 3. Providing explanations regarding AC rating procedures and rating dimensions will increase self-efficacy across AC exercises.

However, perhaps explanations have a greater effect on reactions to some rated exercises than others. Perhaps the exercises typically rated according to dimensions viewed as less objective, such as oral communication, benefit more from explanations. We (Roch et al., 2014) found that the dimension of written communication was rated by participants as more objective than oral communication. Also, perhaps some types of explanations have a greater impact. Two interventions based on explanations will be explored: one focusing on explanations for rating procedures and another focusing on explanations regarding rating dimensions. Both interventions should relate to informational justice.

Recorded AC exercises

Standards for ACs were first endorsed in 1975 (Rupp et al., 2015). Not surprisingly, the standards have changed over the years. The 2000 standards stated: "A systematic procedure must be used by assessors to record specific behavioral observations accurately at the time of observation" (Joiner, 2000, p. 322). However, a major change occurred between the 2000 standards and later versions: AC exercises can now be recorded and rated later (Rupp et al., 2015).

Many ACs are now conducted online. According to Lanik (2019), a 2018 Mercer report suggests that 66% percent of North American companies have online ACs. Even recent scoring innovations for in-person ACs call for videotaped exercises (e.g., Oostrom et al., 2019). Is this movement away from real time ratings viewed positively? Given the lack of previous research or theory regarding this issue, no hypotheses are proposed, but the timing of assessor observation and rating will be investigated.

Research Question 2. Do perceptions of AC exercises differ depending on whether the exercises are rated live and in person versus later and based on video recordings?

Method

Participants

Four hundred and forty-one participants recruited from Amazon Mechanical Turk completed a survey created using Qualtrics that presented in-depth AC details. However, 123 participants did not correctly respond to three random responder checks, and 32 completed the survey in less than half the median time (625 seconds or less versus 1272 seconds), making it unlikely that they carefully read the AC exercise descriptions. The data from these participants were not used, resulting in 286 participants. Four additional attention check items focused on instructions given

to the participants. After a close inspection of these items, wording problems were identified, and these items were not used to screen participants. More detailed explanations are available in the online supplemental material available via the Open Science Framework (OSF, https://osf.io/uqz32/?view_only=af6f4209f12b48d997566651808bb1f7).

The majority were men (58%), Caucasian (77%), and currently worked between 31 and 40 hours per week (57%), with another 28% working more than 40 hours. Their ages varied with the largest categories consisting of between 23 and 30 years (32%) and between 31 and 40 years (46%). A majority (65%) had managerial experience. Fifty-five percent had worked in sales, which was the context for the AC. Twenty-two percent (63 participants) had previously participated in an AC.

Measures and materials

AC overview

Materials from an existing in-person AC (Roch, 2019) were presented to the participants. Participants were asked to put themselves into the shoes of an employee of ACME Phones Unlimited wishing to receive a promotion. All experimental instructions are available via the OSF link provided earlier. The six AC exercises included two leaderless group discussions (LGDs), one focusing on participants deciding fictitious employee bonuses as a group (managerial simulation LGD) and one focusing on participants putting pages of a book titled Zoom (Banyai, 1998) in order as a group (Zoom LGD). An oral presentation (explain qualifications for the position), written case analysis (how to handle an angry customer), personality assessment, and an ability test (cognitive ability test but not identified as such) were also included. The two LGDs, oral presentation, and written case analysis represented rated exercises. Even though cognitive ability tests and personality assessments are not popular assessment center exercises, they have been used in some assessment centers (Thornton & Rupp, 2006).

Experimental manipulations

Participants were given additional instructions relevant to their experimental condition. Two interventions were designed to manipulate informational justice using explanations: one focused on rating dimensions and the other focused on rating procedures. The third intervention focused on whether participants would be rated “live” with assessors in the room as they completed each AC exercise or their exercise performance would be recorded and rated at a later time. Thus, the experimental design consisted of a 2 (rating procedure information, no versus yes) \times 2 (rating dimension information, no versus yes) \times 2 (rating time, real time versus later) crossed factorial design with random assignment to condition.

Measures

After each exercise description, participants completed the following measures: (1). informational justice measure (coefficient alphas ranging from .77 to .87 across exercises) focusing on the explanations given for the exercise ratings and consisting of six items (three justice and three injustice items) from Colquitt et al. (2015), (2). participant motivation measure (coefficient alphas ranging from .88 to .92) consisting of five items proposed by Hedge and Teachout (2000) that we had adapted in our previous studies (Roch et al., 2008; 2014) to focus on AC exercises, (3). Self-efficacy measure (coefficient alphas ranging from .91 to .94) consisting of Bandura’s (2006) 11 item self-efficacy measure, (4). Perceived rating bias measure (coefficient alphas ranging from .84 to .92) based on two items from Goodson & McGee (1991) focusing on assessor rating subjectivity and bias, (5). Fairness item—“Do you believe that the ratings that you would receive for this task will be fair?”. Matthews et al. (2022) showed that one-item justice measures have acceptable

Table 1. Means, Standard Deviations, and Correlations among Study Variables Collapsed Across Assessment Center Exercises and Experimental Interventions

Variable	Mean	SD	1	2	3	4	5
1. Informational Justice	5.45	1.02	–				
2. Motivation	5.87	.91	.65*	–			
3. Bias	3.92	1.28	–.46*	–.21*	–		
4. Self-Efficacy	7.57	1.81	.22*	.43*	.04	–	
5. Fairness	5.36	1.00	.53*	.67	–.23*	.43*	–
6. AC Experience	1.22	.41	–.21*	–.09	.22*	.11	.09

Note: *N* is 285 or 286. Assessment center experience (AC Experience): 1 = No and 2 = Yes. Reliability information for the measures can be found in the Method section.

validity. Justice and fairness are closely related. All items were assessed using seven-point Likert type scales with higher number indicating a greater extent, except for self-efficacy, which was assessed using a 11-point measure ranging from 0% to 100% confidence. Afterward, participants answered demographic questions, including a question regarding whether they had previously attended an assessment center (no or yes).

Results

All analyses were conducted using JASP 17.1 (JASP Team, 2023). Table 1 presents the variable descriptives collapsed across exercises and interventions, including the dichotomous AC experience variable. AC experience was included in all analyses as a between participant variable given concerns that those who experience a selection method may not respond the same as those who had not experienced it (e.g., König et al., 2015). By including AC experience as a between participant variable, any effects found for experience could be further investigated. According to a post hoc power analysis using GPower 3.1.9.7, given 286 participants answering six repeated measures with an average correlation between repeated measures of .70 (which is the case for informational justice), and a medium effect size ($f = .25$), power is .96 if only the 3 interventions are considered (8 conditions) and .89 if AC experience is considered (16 conditions).

Manipulation check—informational justice

Given that the two explanation interventions were designed to influence informational justice perceptions, a 2 (dimension info. intervention) \times 2 (procedure info. intervention) \times 2 (rating time intervention) \times 2 (AC experience) repeated measures ANOVA with the six AC exercise informational justice measures as the repeated measure was conducted. This analysis revealed only one significant within participant effect, a main effect for type of exercise, $F(4.63, 1248.64) = 9.69$, $p < .001$, using the Greenhouse–Geisser correction given that the sphericity assumption was violated. Despite expectations, the ANOVA also revealed only one significant between participants main effect, AC experience, $F(1, 269) = 12.68$, $p < .001$, with those with no AC experience indicating higher levels of informational justice. Out of the 11 possible between participant interactions, only one was significant, a three-way interaction between the interventions, $F(1, 269) = 4.85$, $p = .03$. Follow-up pairwise comparisons using the Bonferroni correction show that the dimension info. intervention had its intended effects only if no procedural info. was given and only in the later rating condition. Thus, there is no indication that the two explanation interventions had their intended effects. See Table 2 for all between participant results for informational justice.

Table 2. Informational Justice ANOVA Results—Between Participants Only

Source	df	<i>F</i>	<i>p</i>
Between Participants Effects			
Proc. Explain	1	0.12	0.73
Dim. Explain	1	0.01	0.92
Rating Time	1	2.99	0.08
AC Exp.	1	12.68	<.001
Proc. Explain * Dim. Explain	1	1.33	0.25
Proc. Explain * Rating Time	1	0.06	0.81
Dim. Explain * Rating Time	1	0.36	0.55
Proc. Explain * AC Exp.	1	9.63×10^{-5}	0.99
Dim. Explain * AC Exp.	1	5.95×10^{-3}	0.94
Rating Time * AC Exp.	1	3.60	0.06
Proc. Explain * Dim. Explain * Rating Time	1	4.85	0.03
Proc. Explain * Dim. Explain * AC Exp.	1	0.36	0.55
Proc. Explain * Rating Time * AC Exp.	1	0.13	0.72
Dim. Explain * Rating Time * AC Exp.	1	9.89×10^{-4}	0.97
Four-way interaction	1	1.55	0.21
Error	269		

Note: Proc. Explain represents the procedural information intervention, an informational justice intervention. Dim. Explain represents the rating dimension information intervention, also an informational justice intervention. Rating time represents the rating timing intervention (live versus later). AC Exp. represents assessment center experience.

Investigation of hypotheses

Even though neither explanation intervention influenced informational justice as expected, all hypotheses were nonetheless explored based on four 2 (dimension info. intervention) \times 2 (procedure info. intervention) \times 2 (rating time intervention) \times 2 (AC experience) repeated measures ANOVAs, one for each DV specified in the hypotheses. Full ANOVA results can be found via the OSF link provided earlier.

Intervention results (between participant results)

Regarding both **motivation** and **perceived bias**, the three interventions and their interactions were nonsignificant. However, a rating timing intervention main effect was found for **fairness**, $F(1, 260) = 4.65$, $p = .03$, suggesting that participants in the later rating condition viewed the exercises as fairer than those in the real time condition. In regard to **self-efficacy**, the interaction between the rating timing intervention, AC experience, and the exercise repeated measure was significant, $F(4.57, 1228.87) = 3.31$, $p = .007$. Follow-up analyses showed that for those with no AC experience, only the exercise repeated measure was significant but for those with AC experience, the interaction between type of exercise and the rating timing intervention was significant, $F(4.32, 233.29) = 3.62$, $p = .006$. Based on post hoc analyses with the Bonferroni correction, this interaction suggests that those with AC experience did not significantly differentiate between the exercises in the real-time rating condition but did so in the later rating condition. In summary, contrary to Hypotheses 2 and 3, the two explanation interventions had no

Table 3. Reactions according to AC Exercise Collapsed Across Experimental Interventions

Exercise	Fairness		Motivation		Bias – No AC Exp.		Bias – AC Exp.	
	M	SD	M	SD	M	SD	M	SD
Ability test	6.02 ^a	1.41	6.12 ^a	1.07	2.30 ^a	1.75	4.04 ^a	2.13
Written case analysis	5.50 ^b	1.50	5.44 ^c	.85	3.81 ^b	1.71	4.62 ^b	1.60
LGD – Mana.	5.17 ^{cd}	1.33	6.05 ^a	.98	4.39 ^c	1.50	4.47 ^{ab}	1.27
LGD – Zoom	5.20 ^{cd}	1.35	5.82 ^b	1.20	3.65 ^b	1.76	4.43 ^{ab}	1.59
Oral presentation	5.23 ^c	1.33	5.97 ^a	1.03	4.24 ^c	1.53	4.68 ^b	1.43
Personality	4.93 ^d	1.60	5.79 ^b	1.16	4.24 ^c	1.74	4.61 ^b	1.64

Note: $n = 285$ or 286 . LGD—Leaderless Group Discussion. LGD Mana. Simulation represents the management simulation LGD. All contrasts used the Bonferroni correction for multiple contrasts and represent contrasts within reactions and not across reactions. Significant differences between exercises indicated with different letter subscripts. For No AC Experience (No AC Exp.), $n = 223$ but for AC Experience (AC Exp.), $n = 62$.

significant effects on participant reactions. However, the rating timing intervention did influence both fairness perceptions and self-efficacy.

AC exercise difference (within participants results)

The AC exercise repeated measure was significant for all reactions, $F(4.05, 1090.64) = 30.92, p < .001$ for **motivation**, $F(4.25, 1139.98) = 32.40, p < .001$ for **bias**, $F(4.57, 1228.87) = 7.11, p < .001$ for **self-efficacy**, and $F(4.57, 1187.35) = 16.79, p < .001$ for **fairness**, all with the Greenhouse–Geisser correction. Two main effects were qualified: An interaction between the exercise repeated measure and AC experience for **bias**, $F(4.25, 1139.98) = 10.80, p < .001$, and the three-way interaction for **self-efficacy** described above.

Table 3 presents the means and standard deviations for **motivation**, **fairness**, and **bias** across experimental conditions, given the lack of significant interactions between the exercise repeated measure and the interventions. The results are presented according to AC experience for **bias** given the significant interaction between the exercise repeated measure and AC experience. Follow-up analyses suggest that those without AC experience differentiated more between the exercises than those with experience, but it should be noted that sample size for those without experience ($n = 62$) is smaller than those with experience ($n = 223$). However, it is notable that the range of means for bias across AC exercises is also greater for those without AC experience than those with experience and that those without AC experience appear to view the ability test as having less bias than those with AC experience.

Table 4 presents the means for **self-efficacy** as a function of both AC experience and rating timing condition, given the significant three-way interaction. As mentioned, unlike those without AC experience, those with AC experience differentiated more between the exercises in the later versus real-time rating condition. However, these results are based on a small n . All participants, except those with AC experience in the real-time rating condition, rated the ability test as the most efficacious exercise, with the written case analysis in second place.

In support of Hypotheses 1a and 1b, the ability test was viewed as significantly fairer and less biased, at least among those with no AC experience, than the other exercises. Those with AC experience did not significantly differentiate between the ability test and both LGDs in terms of bias but did view the ability test as less biased than the other exercises. Contrary to Hypotheses 1c and 1d, surprisingly, the ability test was viewed as one of the most motivating and most self-efficacious exercises.

Table 4. Self-Efficacy as a Function of Assessment Center Experience and Rating Timing Intervention

Exercise	Live rating condition				Later rating condition			
	No AC experience		AC experience		No AC experience		AC experience	
	M	SD	M	SD	M	SD	M	SD
Ability test	7.86	2.19	7.73	1.81	7.89	2.22	8.82	1.81
Written case analysis	7.58	2.25	7.57	2.07	7.88	1.98	8.21	1.81
LGD – Mana.	7.19	1.92	7.53	1.89	7.60	1.95	8.10	1.74
Oral presentation	7.11	2.15	7.70	1.46	7.43	2.07	8.00	2.12
Personality	7.12	2.21	7.94	1.72	7.29	2.01	7.98	1.95
LGD – Zoom	7.06	2.15	7.96	1.62	7.57	2.01	7.90	1.92

Note: For participants in the live rating condition, $n = 103$ for no experience and $n = 33$ for experience. For participants in the later rating condition, $n = 120$ for no experience and $n = 29$ for experience.

In terms of the exploratory question regarding differences among the rated exercises, the written case analysis was viewed as significantly fairer than the other rated exercises but also the least motivating. It was also rated among the least biased exercises and among the most efficacious. It should be noted that the written case analysis differed from the other rated exercises in that there was no oral component (versus oral presentation and LGDs) and required a written response.

Discussion

The results show that 1). AC exercises can be associated with varying levels of reactions, 2). whether exercises are rated live or later matters in terms of fairness perceptions and self-efficacy, and 3). AC experience relates to bias perceptions and self-efficacy. These findings have both theoretical and practical implications.

AC exercise differences

Participants viewed the ability test and written case analysis as both the fairest and least biased of the AC exercises presented to them. Surprisingly, self-efficacy perceptions also tended to be among the highest for the ability test and written case analysis, both relatively objectively assessed exercises. We (Roch et al., 2014) had suggested that individuals may have higher expectations that if they put forth effort, they could perform well in less objectively assessed exercises versus more objectively assessed ones, which appeared to not be the case in this study, assuming that effort-performance expectancy and self-efficacy are closely related (Locke et al., 1986). However, perhaps participants had an easier time thinking of performance criteria for the ability test and written case analysis than for the oral exercises, and the greater uncertainty regarding performance criteria translated into less self-efficacy for the oral exercises.

Being positively perceived in terms of fairness, self-efficacy, and bias did not automatically translate into an exercise being viewed as motivating. The three most motivating exercises were the ability test, oral presentation, and managerial simulation LGD. The oral presentation and the managerial simulation LGD were among the exercises viewed as the least fair, most biased, and approached with the least self-efficacy. Perhaps the oral presentation and managerial simulation LGD benefited from participants believing that putting forth effort will be beneficial to their performance, more so than in the other exercises. Even though participants did not feel very

efficacious regarding their performance in the oral presentation and managerial simulation LGD, they did feel motivated. Further research is needed to determine why this was the case.

The largest discrepancy between the current results and our previous research (Roch et al., 2008; 2014) was regarding the ability test. Similar to the results of our previous studies, participants viewed this exercise as the fairest. However, contrary to our previous research, participants also considered the ability test as among the most motivating and self-efficacious exercises. Participants in our previous studies completed the perception questionnaire after completing the exercises. Perhaps results would be similar across studies if we had assessed reactions before participants completed the exercises.

The findings of this study and our previous studies may shed some light on the AC construct validity problem. Simply put, the construct validity problem is that it is unknown why the OAR (overall assessment rating based on dimension ratings across exercises) has predictive validity given that dimension ratings only modestly correlate across AC exercises (e.g., Hoffman et al., 2015). Lance (2008) and Lievens (2002) suggest that inconsistent cross-situational participant behaviors contribute to the AC construct validity problem. Perhaps inconsistent cross-situational participant behavior may at least partly be a function of different participant reactions to specific AC exercises.

Interventions

The two explanation-based interventions did not have their anticipated effects, not unlike previous research suggesting that explanation-based interventions have smaller effects in scenario/simulation-based research in comparison to research in selection contexts (Truxillo et al., 2009). The only specific selection methods investigated in the Truxillo et al. (2009) meta-analysis were cognitive ability tests and personality assessments. As mentioned earlier, it is possible that explanations may have a greater effect for rated exercises than objectively scored ones given the greater ambiguity regarding ratings versus scoring an exercise based on verifiably correct answers, but our results suggest that this may not be the case. However, the current study did not have adequate power to investigate small effect sizes.

According to Truxillo et al. (2009), explanations can be viewed as focusing on structural fairness (as in the current study and in most previous research) or social fairness, focusing on interpersonal sensitivity. When explaining the lack of success for their structural explanation intervention, Melchers and Körner (2019) suggest that structural explanations increase anxiety and that perhaps explanation-based interventions focusing on social fairness may be more successful.

However, the exploratory rating timing intervention did appear to matter. This intervention had a significant main effect on fairness perceptions, with participants in the later rating condition reporting higher fairness levels across AC exercises than those in the live rating condition. However, the relationship was more complex regarding self-efficacy, given the three-way interaction between AC experience, the rating timing intervention, and the exercise repeated measure. For those with no AC experience, only the exercise repeated measure was significant. However, the rating timing intervention did matter for those with AC experience. In the live rating condition, participants did not differentiate between exercises in regard to self-efficacy but in the later condition, they did differentiate. Those with AC experience also appeared to differentiate less among the exercises in terms of bias than those with no AC experience. However, these results regarding AC experience are tentative given that only 63 participants had AC experience. More research is needed to explore why those with no AC experience tend to differentiate among AC exercises more than those with AC experience.

Practical implications

Given the fairness perception benefit associated with the belief that exercise performance would be recorded and later rated, online ACs have the advantage of having this benefit implicit. If an organization is concerned about justice/fairness perceptions, which have been shown to relate to a host of organizational attitudes and behaviors (Colquitt et al., 2013) in addition to acceptance of job offers (Harold et al., 2016), AC exercises should be recorded and rated at a later time.

Also, even though the differences between participants with and without AC experience were not extensive, AC experience did relate to bias and self-efficacy perceptions. Organizations should be aware that those without AC experience appear to distinguish more between AC exercises in terms of bias, which may relate to greater cross-situational behavior inconsistencies in comparison with those with AC experience. Furthermore, given that those with AC experience tended to view AC exercises generally as more biased, they may be more unlikely to accept the AC feedback reports than those without AC experience.

Potential limitations and future research

As with all research, this study has potential limitations that provide avenues for future research. The participants did not complete the exercises but were given in-depth exercise descriptions and asked to imagine themselves completing the exercises motivated to perform well. It would be beneficial for future researchers to assess perceptions of AC exercises before and after exercise completion, perhaps shedding light on discrepancies between the current results and our previous research investigating AC exercises (Roch et al., 2008; 2014).

Furthermore, in the current study, previous AC experience was assessed as a dichotomous variable—no or yes. It would be helpful if future researchers had more information regarding the recency and amount of previous AC experience, along with not only a larger sample size but also one more evenly divided between those with and without AC experience to detect differences with smaller effect sizes. Nevertheless, it appears that this study is the first to empirically investigate whether those with AC experience respond differently than those without AC experiences to specific AC exercises.

In summary, the current study shows that it is possible for an AC exercise to be perceived as fair, not biased, but still motivating and self-efficacious. However, most common AC exercises are not perceived as such. Future researchers should further investigate what design features contribute to this desirable conglomeration of features. Also, it appears that previous AC experience does relate to bias perceptions and self-efficacy, preliminary findings worthy of additional research. Finally, it appears that AC exercises are viewed as fairer if they are recorded and rated later; another topic worthy of further research to better understand why this is the case.

Acknowledgments. Funds for this project were provided from a Douglas W. Bray and Ann Howard Research Grant from the SIOP Foundation.

References

- Ambrose, M. L., & Schminke, M. (2009). The role of overall justice judgments in organizational justice research: A test of mediation. *Journal of Applied Psychology*, 94(2), 491–500. <https://doi.org/10.1037/a0013203>.
- Anderson, N., & Witvliet, C. (2008). Fairness reactions to personnel selection methods: An international comparison between the Netherlands, the United States, France, Spain, Portugal, and Singapore. *International Journal of Selection and Assessment*, 16(1), 1–13. <https://doi.org/10.1111/j.1468-2389.2008.00404.x>.
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37, 122–147.
- Banyai, I. (1998). *Zoom*. Puffin Books.
- Bell, B. S., Wiechmann, D., & Ryan, A. M. (2006). Consequences of organizational justice expectations in a selection system. *Journal of Applied Psychology*, 91, 455–466. <https://doi.org/10.1037/0021-9010.91.2.455>.

- Chan, D., Schmitt, N., DeShon, R. P., Clause, C. S., & Delbridge, K. (1997). Reactions to cognitive ability tests: The relationships between race, test performance, face validity perceptions, and test-taking motivation. *Journal of Applied Psychology*, *82*, 300–310. <https://doi.org/10.1037/0021-9010.82.2.300>.
- Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology*, *86*, 386–400. <https://doi.org/10.1037/0021-9010.86.3.386>.
- Colquitt, J. A., Long, D. M., Rodell, J., & Halvorsen-Ganepola, M. (2015). Adding the “in” to justice: A qualitative and quantitative investigation of the differential effects of justice rule adherence and violation. *Journal of Applied Psychology*, *100*, 278–297. <https://doi.org/10.1037/a0038131>.
- Colquitt, J. A., Scott, B., Rodell, J., Long, D., Zapata, C., Conlon, D., & Wesson, W. (2013). Justice at the millennium, a decade later: A meta-analytic test of social exchange and affect-based perspectives. *Journal of Applied Psychology*, *98*, 199–236. <https://doi.org/10.1037/a0031757>.
- Goodson, J. R., & McGee, G. W. (1991). Enhancing individual perceptions of objectivity in performance appraisal. *Journal of Business Research*, *22*(4), 293–303. [https://doi.org/10.1016/0148-2963\(91\)90036-W](https://doi.org/10.1016/0148-2963(91)90036-W).
- Harold, C. M., Holtz, B. C., Griepentrog, B. K., Brewer, L. M., & Marsh, S. M. (2016). Investigating the effects of applicant justice perceptions on job offer acceptance. *Personnel Psychology*, *69*(1), 199–227. <https://doi.org/10.1111/peps.12101>.
- Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, *56*, 639–683. <https://doi.org/10.1111/j.1744-6570.2004.00003.x>.
- Hedge, J. W., & Teachout, M. S. (2000). Exploring the concept of acceptability as a criterion for evaluation performance measures. *Group and Organization Management*, *25*, 22–44. <https://doi.org/10.1177/1059601100251003>.
- Hoffman, B. J., Kennedy, C. L., LoPilato A. C., Monahan, E. L., & Lance, C. E. (2015). A review of the content, criterion-related, and construct-related validity of AC exercises. *Journal of Applied Psychology*, *100*, 1143–1168. <https://doi.org/10.1037/a0038707>.
- Hülsheger, U. R., & Anderson, N. (2009). Applicant perspectives in selection: Going beyond preference reactions. *International Journal of Selection and Assessment*, *17*(4), 335–345. <https://doi.org/10.1111/j.1468-2389.2009.00477.x>.
- Joiner, D. A. (2000). Guidelines and ethical considerations for AC operations: International task force on AC guidelines. *Public Personnel Management*, *29*(3), 315–332. <https://doi.org/10.1177/009102600002900302>.
- König, C. J., Fell, C. B., Steffen, V., & Vanderveken, S. (2015). Applicant reactions are similar across countries: A refined replication with AC data from the European Union. *Journal of Personnel Psychology*, *14*(4), 213–217. <https://doi.org/10.1027/1866-5888/a000142>.
- Krause, D. E., & Thornton, G. C. (2009). A cross-cultural look at AC practices: Survey results from Western Europe and North America. *Journal of Applied Psychology: An International Review*, *58*, 557–585. <https://doi.org/10.1111/j.1464-0597.2008.00371.x>.
- Kravitz, D. A., Stinson, V., & Chavez, T. L. (1996). Evaluations of tests used for making selection and promotion decisions. *International Journal of Selection and Assessment*, *4*(1), 24–34. <https://doi.org/10.1111/j.1468-2389.1996.tb00045.x>.
- Lance, C. E. (2008). Why ACs do not work the way they are supposed to. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *1*, 84–97. <https://doi.org/10.1111/j.1754-9434.2007.00017.x>.
- Lanik, M. Why use assessment centers? Because they increase quality of hires, *Pinsight*. <https://www.pinsight.com/blog/why-use-assessment-centers/>.
- Leventhal, G. S. (1980). What should be done with equity theory? New approaches to the study of fairness in social relationships. In K. Gergen, M. Greenberg, & R. Willis (Eds.), *Social exchange: Advances in theory and research* (pp. 27–55). Plenum Press.
- Lievens, F. (2002). Trying to understand the different pieces of the construct validity puzzle of ACs: An examination of assessor and assessee effects. *Journal of Applied Psychology*, *87*, 675–686. <https://doi.org/10.1037/0021-9010.87.4.675>.
- Locke, E. A., Motowidlo, S. J., & Bobko, P. (1986). Using self-efficacy theory to resolve the conflict between goal-setting theory and expectancy theory in organizational behavior and industrial/organizational psychology. *Journal of Social and Clinical Psychology*, *4*(3), 328–338. <https://doi.org/10.1521/jscp.1986.4.3.328>.
- Macan, T. H., Avedon, M. J., Paese, M., & Smith, D. E. (1994). The effects of applicants’ reactions to cognitive ability tests and an AC. *Personnel Psychology*, *47*(4), 715–738. <https://doi.org/10.1111/j.1744-6570.1994.tb01573.x>.
- Matthews, R. A., Pineault, L., & Hong, Y. H. (2022). Normalizing the use of single-item measures: Validation of the single-item compendium for organizational psychology. *Journal of Business and Psychology*, *37*(4), 639–673. <https://doi.org/10.1007/s10869-022-09813-3>.
- Maurer, T. J., Solamon, J. M., Andrews, K. D., & Troxtel, D. D. (2001). Interviewee coaching, preparation strategies, and response strategies in relation to performance in situational employment interviews: An extension of Maurer, Solamon, and Troxtel, 1998. *Journal of Applied Psychology*, *86*(4), 709–717. <https://doi.org/10.1037/0021-9010.86.4.709>.
- Melchers, K. G., & Körner, B. (2019). Is it possible to improve test takers’ perceptions of ability tests by providing an explanation? *Journal of Personnel Psychology*, *18*(1), 1–9. <https://doi.org/10.1027/1866-5888/a000212>.
- Merkulova, N., Melchers, K. G., Kleinmann, M., Annen, H., & Tresch, T. S. (2014). Effects of individual differences on applicant perceptions of an operational AC. *International Journal of Selection and Assessment*, *22*, 355–370.

- Moscoso, S., & Salgado, J. F. (2004). Fairness reactions to personnel selection techniques in Spain and Portugal. *International Journal of Selection and Assessment*, 12(1-2), 187–196. <https://doi.org/10.1111/j.0965-075X.2004.00273.x>.
- Oostrom, J. K., Lehmann-Willenbrock, N., & Klehe, U. (2019). A new scoring procedure in ACs: insights from interaction analysis Personnel. *Assessment and Decisions*, 5(1). <https://doi.org/10.25035/pad.2019.01.005>.
- Phillips, J. M., & Gully, S. M. (2002). Fairness reactions to personnel selection techniques in Singapore and the United States. *The International Journal of Human Resource Management*, 13(8), 1186–1205. <https://doi.org/10.1080/09585190210149475>.
- Ployhart, R. E., & Ehrhart, M. G. (2002). Modeling the practical effects of applicant reactions: Subgroup differences in test-taking motivation, test performance, and selection rates. *International Journal of Selection and Assessment*, 10(4), 258–270. <https://doi.org/10.1111/1468-2389.00216>.
- Rego, A., Sousa, F., Marques, C., Pina, M., & Cunha, E. (2012). Retail employees' self-efficacy and hope predicting their positive affect and creativity. *European Journal of Work and Organizational Psychology*, 21(6), 923–945. <https://doi.org/10.1080/1359432X.2011.610891>.
- Roch, S. G. Applicant perceptions of assessment center exercise scoring objectivity: motivation and justice 2019. Unpublished manuscript.
- Roch, S. G., Mishra, V., & Trombini, E. (2014). Does selection measure scoring influence motivation: One size fits all? *International Journal of Selection and Assessment*, 22, 23–38. <https://doi.org/10.1111/ijsa.12054>.
- Roch, S. G., Trombini, E., & Mishra, V. Rater teams, perceived dimension subjectivity, and AC participant motivation. In: *23rd Annual Meeting of the Society for Industrial and Organizational Psychology*, San Francisco, CA, 2008.
- Rupp, D. E., Gibbons, A. M., Baldwin, A. M., Snyder, L. A., Spain, S. M., Woo, S. E., Brummel, B. J., Sims, C. A., & Kim, M. (2006). An initial validation of developmental ACs as accurate assessments and effective training interventions. *The Psychologist-Manager Journal*, 9, 171–200. https://doi.org/10.1207/s15503461tpmj0902_7.
- Rupp, D. E., Hoffman, B. J., Bischof, D., Byham, W., Collins, L., Gibbons, A., & Thornton, G. (2015). Guidelines and ethical considerations for AC operations. *Journal of Management*, 41(4), 1244–1273. <https://doi.org/10.1177/0149206314567780>.
- Sanchez, R. J., Truxillo, D. M., & Bauer, T. N. (2000). Development and examination of an expectancy-based measure of test-taking motivation. *Journal of Applied Psychology*, 85(5), 739–750. <https://psycnet.apa.org/doi/10.1037/0021-9010.85.5.739>
- Smither, J. W., Reilly, R. R., Millsap, R. E., Pearlman, K., & Stoffey, R. W. (1993). Applicant reactions to selection procedures. *Personnel Psychology*, 46(1), 49–76. <https://doi.org/10.1111/j.1744-6570.1993.tb00867.x>.
- Snyder, J., & Shahani-Denning, C. (2012). Fairness reactions to personnel selection methods: A look at professionals in Mumbai, India. *International Journal of Selection and Assessment*, 20(3), 297–307. <https://doi.org/10.1111/j.1468-2389.2012.00601.x>.
- Steiner, D. D., & Gilliland, S. W. (1996). Fairness reactions to personnel selection techniques in France and the United States. *Journal of Applied Psychology*, 81(2), 134–141. <https://doi.org/10.1037/0021-9010.81.2.134>.
- JASP Team (2003). *JASP (Version 0.17.3) [Computer software]*.
- Thornton, G. C., & Rupp, D. (2006). *ACs in human resource management: Strategies for prediction, diagnosis, and development*. Lawrence Erlbaum Associates.
- Truxillo, D. M., Bodner, T. E., Bertolino, M., T. N., Bauer, & Yonce, C. A. (2009). Effects of explanations on applicant reactions: A meta-analytic review. *International Journal of Selection and Assessment*, 17, 346–361. <https://doi.org/10.1111/j.1468-2389.2009.00478.x>.
- Zapata-Phelan, C. P., Colquitt, J. A., Scott, B., & Livingston, B. (2009). Procedural justice, interactional justice, and task performance: The mediating role of intrinsic motivation. *Organizational Behavior and Human Decision Processes*, 108, 93–105. <https://doi.org/10.1016/j.obhdp.2008.08.001>.