Analysis of a situational judgment test for teamwork as a preselection tool for an assessment center: A construct-based approach

Goerke, Panja; Maier, Julia

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:
Goerke, P., & Maier, J. (2022). Analysis of a situational judgment test for teamwork as a preselection tool for an assessment center: A construct-based approach. *International Journal of Selection and Assessment*, *30*(3), 456-464. https://doi.org/10.1111/ijsa.12391

Received: 24 September 2021 | Revised: 30 May 2022 | Accepted: 1 June 2022

DOI: 10.1111/ijsa.12391

SHORT COMMUNICATION

Information Exchange Article

International Journal of Selection and Assessment | WILEY

Analysis of a situational judgment test for teamwork as a preselection tool for an assessment center: A construct-based approach

Panja Goerke | Julia Maier

Department of Aviation and Space Psychology, DLR-German Aerospace Center, Hamburg, Germany

Correspondence
Panja Goerke, Department of Aviation and Space Psychology, DLR-German Aerospace Center, Sportallee 54a, 22335 Hamburg, Germany.
Email: panja.goerke@dlr.de

Abstract

Situational judgment tests (SJTs) constitute an efficient and well-established method for predicting professional success. However, more information is needed regarding their relationship to applicants' behavior as observed in an assessment center (AC) at the construct level. In the present study, an SJT for teamwork (SJT-TW) was assessed in terms of its suitability as a preselection tool for ACs in the context of a multistage selection. Data were collected from 276 applicants for ab initio pilot training during their selection process. Results from the regression analyses showed that the SJT-TW test performance is a significant predictor for the AC result and it contributed more than knowledge, cognitive performance tests, and personality scales thereto. The SJT-TW also caused significant increments for the prediction of single AC dimensions. Therefore, it was concluded that construct-based SJTs are useful for the assessment of behavioral constructs and can complement selection processes as a preselection tool.

KEYWORDS

assessment center, preselection, situational judgment test, teamwork

Practitioner points

- Situational judgment tests constitute a well-established and efficient selection method for assessing social competence and predicting job success.
- Relatively little is known concerning the relationship between construct-based situational judgment tests and applicants' behavior as observed in an assessment center.
- Our results confirm that a situational judgment test for teamwork can be used to predict applicants' behavior.
- Construct-based situational judgment tests as preselection tools offer the possibility of efficiently, objectively, and flexibly assessing behavioral aspects.

1 | INTRODUCTION

Situational judgment tests (SJTs) have gained increasing attention in personnel selection and research in the last decades (e.g., Wheekley et al., 2015). Like assessment centers (ACs) and work samples, SJTs can be classified as simulation-based instruments (Lievens & De Soete, 2015), as they confront candidates with realistic job situations. Applicants work on written or video-based descriptions of job-relevant situations focussing on a dilemma or problem and are required to respond by choosing from different predefined behavioral reactions (e.g., McDaniel et al., 2007; Wheekley et al., 2015). Compared to ACs, which are high-fidelity simulations, SJTs are low-fidelity simulations as they "only" assess procedural knowledge and behavioral intentions (Thornton & Rupp, 2006). Depending on the instruction format ("would do" vs. "should do"), either a behavioral or a knowledge-based answer is elicited (McDaniel & Nguygen, 2001). The behavioral instruction focuses on the action that the test taker would most likely perform and is associated with typical performance, whereas the knowledge instruction asks for the best possible answer and is expected to evoke a maximum performance response (McDaniel & Nguyen, 2001; McDaniel et al., 2007). In their meta-analysis, McDaniel et al. (2007) showed that SJTs with knowledge instructions correlate higher with cognitive ability, while SJTs with behavioral instructions correlate higher with personality. Furthermore, SJTs with knowledge instructions were less prone to faking in (high-stakes) selection contexts (Nguyen et al., 2005; Zhang et al., 2021) due to their focus on maximum performance.

Revisiting meta-analytical validities in personnel selection, Sackett et al. (2021) showed that SJTs are valid predictors of job performance, with operational validities that are comparable to those of ACs, based on several meta-analytical studies. They reported an operational validity of 0.26 for SJTs and an average validity estimate across different meta-analyses of 0.29 for ACs. Thus, SJTs as low-fidelity simulations have the potential to compete with high-fidelity simulations when trying to predict job-related behavior. Moreover, as McDaniel et al. (2007) reported in their meta-analysis, SJTs have meaningful incremental validity for job performance over Big Five personality traits and modest incremental validity over cognitive ability.

High-fidelity simulations such as work samples or ACs require applicants to respond to job-related situations with actual behavior (Thornton & Rupp, 2006; Wheekley et al., 2015). As they are rather complex and resource-intensive, they are usually used later in the selection process. Contrarily, a larger group of candidates can simultaneously perform SJTs; SJTs can also be presented digitally and are less costly than high-fidelity tools. Consequently, it would be economically reasonable to use SJTs as a preselection tool in personnel selection contexts (Lievens & De Soete, 2015; Lievens et al., 2021). Taking into account these considerations, the main focus of this paper is to analyze whether an SJT can be used as a preselection tool for an AC in a multistage selection process.

Generally, SJTs are aimed at measuring job-realistic performance by covering a broad range of job-specific situations and a variety of constructs (McDaniel et al., 2001). This often leads to heterogeneous item content and rather low internal consistencies (Lievens et al., 2021). Therefore, a common practice is to aggregate results to an overall SJT score (Christian et al., 2010). However, when analyzing the predictive validity of personnel selection methods, different researchers have stressed the need for considering specific constructs (e.g., Arthur & Villado, 2008; McDaniel et al., 2007; Ployhart, 2006; Roth et al., 2008). This claim can be transferred to SJTs when trying to understand how and why SJTs work in a selection context (Christian et al., 2010). Furthermore, recent approaches have shown that SJTs developed to measure concrete constructs have higher validity than SJTs with heterogeneous composite scores (Christian et al., 2010) and minimize the effects of faking compared to a self-report of personality (Kasten et al., 2018). Specifically, Christian et al. (2010) found that SJTs measuring constructs such as teamwork skills have higher validities for overall job performance, as well as for contextual performance, than SJTs assessing other applied social skills, heterogeneous composites, or job knowledge. They highlighted the importance of a construct-based approach in SJT research for two reasons: It would be insightful for the question of why an SJT is related to a certain criterion or why it is not, and it would allow for a more accurate comparison between different selection methods. Arthur et al. (2003) called for the same approach regarding ACs and suggested focussing on dimension ratings instead of overall scores. In their meta-analysis, they found criterion-oriented validities among clusters of AC dimensions (e.g., "consideration/awareness of others" and "organization and planning") and job-related criteria. Additionally, Lievens and Patterson (2011) analyzed the validity of a construct-based SJT in comparison to an AC and a job knowledge test in high-stakes selection. In the context of a multistage selection process for general practitioners, $N = 196$ physicians took part in the study. While the job knowledge test addressed applied clinical knowledge, the written SJT and the AC were intended to measure the same five performance dimensions: communication, empathy, professional integrity, coping with pressure, and problem-solving. The results showed predictive validity for the criterion of job performance for all three selection measures. The SJT had the highest correlation with overall job performance ($r = .37$), followed by the job knowledge test ($r = .36$) and the AC ($r = .30$), even though there was no significant difference in criterion-related validity between the SJT and the AC. Furthermore, the authors reported that the SJT and AC were substantially correlated ($r = .43$) and that both instruments had incremental validity over the job knowledge test for the prediction of job performance. Nevertheless, their analyses were based on aggregated values, although they originally had collected data on the dimension/construct level for both measures (SJT and AC).

Considering previous findings, in this paper, we analyze the relationship between SJT and AC performance in the context of a multistage selection process; as already mentioned, Lievens and Patterson (2011) reported a substantial correlation between SJT and AC performance. This led to the question of whether observations of applicants' behavior in an AC could be predicted by SJT performance. Thus, the predictive power of an SJT for the outcome of the AC is analyzed to discover more about its level of contribution over

knowledge and cognitive performance tests and personality scales, which were used as a first selection stage in a high-stakes multistage selection procedure. To elaborate on the relationship between SJT and AC performance, we follow the construct-level approach: We use an SJT that was constructed for selecting individuals for teamwork and assess its predictive power for the outcome of the AC and AC performance in teamwork-relevant dimensions. Due to their different diagnostic approaches, a moderate relationship between construct assessments made by SJTs and ACs is expected (Lievens & Patterson, 2011).

2 | MATERIALS AND METHODS

The study took place as part of a multistage selection process for ab initio pilots (Zinn et al., 2020). In the first selection stage, applicants undertook several computer-based cognitive performance and knowledge tests, as well as a personality questionnaire. During the second selection stage, the situational judgment test for teamwork (SJT-TW) was completed subsequent to an AC.

2.1 | Sample

The sample consisted of $N = 276$ applicants who had taken part in both selection stages. Most of the subjects were male (female = 37, male = 239), representing the common gender composition of pilot applicants. The subjects' ages ranged between 18 and 29 years ($M = 21.5$, $SD = 2.9$).

3 | MEASURES

3.1 | Knowledge and cognitive performance tests

The computer-based tests measured different areas of job knowledge, namely English language skills, technics, and mathematics. Moreover, different aspects of general mental ability were assessed: memory, concentration, and spatial orientation. All tests were developed in our institution and were administered and monitored regarding their psychometric properties over a long period during the first stage of pilot selection.

3.2 | Personality questionnaire

The Temperament Structure Scales (TSS; Goeters et al., 1993) comprise a nonclinical personality questionnaire developed for pilot selection in the 1970s. They contain 183 items referring to behavior intentions distributed among 10 dimensions (extraversion, emotional instability, dominance, aggressiveness, rigidity, vitality, empathy, spoiltness, mobility, and achievement motivation) plus a control scale. To achieve comparability with the Big Five, most of the scales were aggregated. Based on principal component factor analysis including subsequent Varimax rotation, scale mean values were averaged to represent three of the Big Five constructs: agreeableness (empathy, dominance [−] and spoiltness [−]), conscientiousness (rigidity and achievement motivation) and neuroticism (emotional instability and aggressiveness). The TSS scale for extraversion was maintained, as it was directly comparable with the corresponding Big Five scale ($r = .84$; Mittelstädt et al., 2016). As the TSSs were developed especially for the context of pilot selection, they were not based on any specific personality theory and, thus, did not include the scale "openness for experience."

3.3 | Assessment center

The AC consisted of role play, which included a dilemma situation requiring the candidate to deal with an upset role player, as well as a computer-assisted group discussion exercise with different phases (planning: e.g., rearrangement of passengers with high demands and medium time pressure; conflict: e.g., group discussions concerning vocational promotion). Trained observers rated the performance of the applicants in each phase on four behavioral dimensions: leadership (specifying goals and making decisions), teamwork/communication (interacting openly and responsibly with others, transferring information), adherence to procedures (applying rules correctly and in a disciplined manner) and resilience (maintaining effective performance and having no stress symptoms) on a six-point rating scale. To pass the AC, candidates had to meet the requirements (pass the cut-off scores) for all four dimensions.

3.4 | Situational judgment test

The construct-based SJT-TW was developed by Gatzka and Volmer (2017) to measure how effectively a person can act as a member of a team or working group. It consists of 12 items covering seven categories of teamwork behaviors (e.g., cooperation, planning and organization, and communication) and was developed for broad applicability in different teamwork contexts. The instruction was knowledge-based. Each item had to be answered by marking the best and the worst behavior option (Figure 1). The scoring was based on expert judgments: For each response option, $N = 109$ experts made an effectiveness rating on a five-point scale. The overall score per item was calculated by subtracting the effectiveness rating for the chosen worst option from the rating for the chosen best option.

4 | PROCEDURE

The whole selection process extended over a longer period: Applicants participated in the first selection stage (duration: 1 day) several weeks before the second selection stage (duration of the whole stage: 2 days) took place. Only those who had passed the first

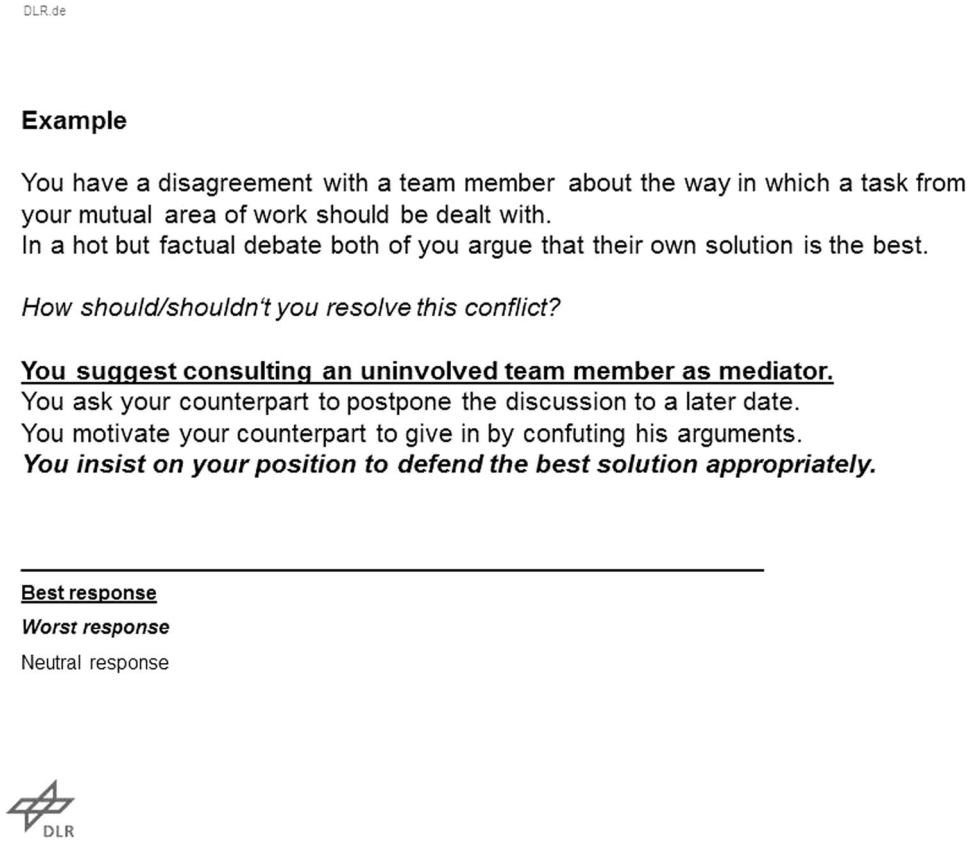

FIGURE 1 Example item for the situational judgment test for teamwork

selection stage were admitted to the second stage. First, candidates had to perform the role play, after which they participated in the group discussion. They then had to fill in the SJT-TW. At the end of the day, candidates received feedback about their performance. Those who passed the AC took part in further selection stages on the following day, but these were not part of the present analyses.

5 | RESULTS

As a first step, bivariate correlations were calculated to analyze the relationships between the SJT-TW and the established selection tools in Stages 1 and 2. In Table 1, the descriptive statistics for knowledge tests, cognitive performance tests, and personality scales are presented, as well as the intercorrelations and correlation coefficients between these instruments and the SJT-TW. No meaningful correlations were found between the SJT-TW and the cognitive performance tests or personality scales. Significant positive relationships of medium height were obtained only with the knowledge tests (English, technics, and mathematics). The results of the correlation analyses between the SJT-TW and AC dimensions are also shown in Table 1. In line with our expectations, medium relationships between SJT-TW performance and observer ratings with most AC dimensions were found. Further, the intercorrelations among AC dimensions were moderately strong.

Second, the contribution of the SJT-TW to the outcome of the AC (pass, $n = 143$ vs. fail, $n = 133$) was analyzed. Candidates who had successfully completed the AC were compared with those who had not. The results of a t test confirmed that candidates who passed the AC performed significantly better on the SJT-TW ($M_{SJT\_ACpass} = 16.98$, $SD = 3.32$) than those who failed, $M_{SJT\_ACfail} = 15.33$, $SD = 3.60$; $t_{(274)} = 3.94$, $p(<.001)$.

To determine the relevance of the SJT-TW in the context of the established tests used in the first selection stage, binary logistic regression using the criterion "AC-result" was calculated. In the first block, knowledge and cognitive performance tests were included; in the second block, personality scales were added; and in the third block, the SJT-TW was incorporated. The results are given in Table 2.

From the overall model evaluation, it can be inferred that Model 3, which included all tests plus the SJT-TW, had the largest predictive power. With Nagelkerke's $R^2 = .20$ and odds ratios < 1.5, it had a medium effect according to Cohen ($f = 0.50$). The SJT-TW had the highest significant regression coefficient, and its addition to the equation led to a significant gain in χ^2. Moreover, significant coefficients were identified for the cognitive ability "memory" and the personality trait "agreeableness."

In addition to this global analysis, hierarchical multiple regression analyses were calculated for each AC dimension (leadership, teamwork/communication, adherence to procedures, and resilience). The focus was on the question, which of the AC dimensions was

TABLE 1 Descriptives and correlations for the knowledge and cognitive performance tests, personality scales, assessment center dimensions and situational judgment test for teamwork (SJT-TW)

	M	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Knowledge tests																	
1. English	6.17	1.47															
2. Technics	6.19	1.53	.12*														
3. Mathematics	6.34	1.22	.21***	.37***													
Cognitive performance tests																	
4. Memory	6.56	1.24	.09	.34***	.31***												
5. Concentration	6.47	1.12	.04	.34***	.25***	.60***											
6. Spatial orientation	6.56	1.26	.09	.29***	.24***	.56***	.56***										
Personality scales																	
7. Extraversion	5.97	2.18	−.12*	−.05	.04	−.07	.00	−.10									
8. Conscientiousness	5.07	1.54	−.02	−.11	.02	.07	.17**	.10	−.01								
9. Agreeableness	4.61	1.40	.02	.05	.05	.10	.13*	.02	.10	.07							
10. Neuroticism	4.91	1.40	−.01	.01	−.01	.02	.02	.01	−.01	−.05	.02						
Assessment center																	
11. Leadership	3.09	0.56	.09	.16**	.20**	.22***	.12*	.11	.17**	−.06	.15*	.04					
12. Teamwork/communication	3.37	0.49	.03	.19**	.11	.28***	.22***	.20**	.02	.07	.08	−.03	.66***				
13. Adherence to procedures	3.54	0.29	.10	.11	.14*	.17**	.08	.07	−.06	−.14*	−.03	−.01	.39***	.42***			
14. Resilience	3.21	0.48	.10	.05	.18**	.16**	.09	.07	.09	−.01	.13*	.03	.67***	.55***	.42***		
15. Pass/fail	0.51	0.50	.00	.11	.07	.25**	.20**	.17**	.01	.05	.14*	−.05	.64***	.65***	.35***	.57***	
16. SJT-TW	16.19	3.55	.12*	.13*	.16**	.05	−.04	.04	.01	−.01	−.11	.03	.26***	.22***	.08	.26***	.23***

Note: Reliability coefficient were as follows: English, Cronbach's $\alpha = .90$; technics: technical comprehension, Cronbach's $\alpha = .78$ and physics, Cronbach's $\alpha = .78$; mathematics: mathematical reasoning, Cronbach's $\alpha = .78$ and mental arithmetic, Cronbach's $\alpha = .84$; memory: running memory span, $r_{tt} = .76$ and visual memory capacity, $r_{tt} = .78$; concentration: visual perception speed, $r_{tt} = .90$ and selective attention, $r_{tt} = .93$; spatial orientation: mental rotation, $r_{tt} = .91$ and spatial relations, Cronbach's $\alpha = .83$; extraversion, Cronbach's $\alpha = .80$; conscientiousness: rigidity, Cronbach's $\alpha = .80$ and achievement motivation, Cronbach's $\alpha = .72$; agreeableness: empathy, Cronbach's $\alpha = .76$, dominance, Cronbach's $\alpha = .81$ and spoiltness, Cronbach's $\alpha = .69$; neuroticism: emotional instability, Cronbach's $\alpha = .79$ and aggressiveness, Cronbach's $\alpha = .80$; SJT-TW: Cronbach's α between .33 and .67 (Gatzka & Volmer, 2017).

*$p < .05$; **$p < .01$; ***$p < .001$.

TABLE 2 Results of the binary logistic regression predicting the probability of the assessment center result (pass/fail)

	B	*SE B*	Wald	Odds ratio (95% CI)	Block 1	Block 2	Block 3
Block 1							
English	−0.09	0.09	0.94	0.91 (0.76–1.10)			
Technics	−0.02	0.10	0.03	0.98 (0.81–1.19)			
Mathematics	−0.12	0.12	0.98	0.88 (0.69–1.13)			
Memory	0.36	0.15	5.94*	1.43 (1.07–1.90)			
Concentration	0.14	0.16	0.78	1.16 (0.84–1.59)			
Spatial orientation	0.06	0.14	0.21	1.06 (0.82–1.39)			
Block 2							
Extraversion	0.02	0.06	0.07	1.02 (0.89–1.15)			
Conscientiousness	0.02	0.09	0.05	1.02 (0.86–1.21)			
Agreeableness	0.32	0.12	6.96**	1.37 (1.09–1.73)			
Neuroticism	−0.11	0.09	1.37	0.90 (0.75–1.08)			
Block 3							
SJT-TW	0.17	0.04	17.82***	1.19 (1.10–1.29)			
Overall model evaluation χ^2 (*df*)					18.70 (6)**	24.39 (10)**	44.30 (11)***
$\Delta\chi^2$ (*df*)					n. a.	5.69 (4)	19.91 (1)***
Nagelkerke's R^2					.09	.11	.20
Goodness of fit ($df = 8$)					5.43 ($p = .711$)	19.58 ($p = .012$)	8.28 ($p = .407$)

Note: Statistical coefficients for individual predictors are given for the model obtained in Block 3.
Abbreviations: CI, confidence interval; n.a., not available; SJT-TW, situational judgment test for teamwork.
*$p < .05$; **$p < .01$; ***$p < .001$.

predicted best by the established tests used in the first selection stage and the SJT-TW. In each analysis, the first predictor block contained knowledge and cognitive performance tests; in the second block, personality was added; and in the third block, the SJT-TW was included. There was no multicollinearity in the data set, and tolerance ($M = 0.78$, $SD = 0.18$) and VIF statistics ($M = 1.36$, $SD = 0.34$) were well within critical limits for all criteria (Hair et al., 2018). The global model parameters for the multiple regressions are reported in Table 3. In summary, comparable levels of variance were explained for the dimensions of leadership ($R^2 = .16$), resilience ($R^2 = .14$) and teamwork/communication ($R^2 = .13$). The SJT-TW caused the largest increment for the dimension of resilience ($\Delta R^2 = .07$), although the effects were comparable for leadership ($\Delta R^2 = .06$) and teamwork/communication ($\Delta R^2 = .04$). However, the SJT-TW did not meaningfully contribute to the dimension of adherence to procedures.

Finally, the predictive power of the SJT-TW for later training performance was analyzed. The only criterion data that was currently available was training progress. From all candidates in our sample who had passed the whole selection process and started flight training ($n = 70$), only three appeared in an extra hearing. Their mean SJT-TW scores were slightly lower than the mean scores of all other candidates with a positive selection result ($M_{\text{hearing}} = 15.50$, $SD = 4.40$, $n = 3$; $M_{\text{no hearing}} = 16.92$, $SD = 3.43$, $n = 67$). However, due to the very small sample, there was no statistically significant difference between the two groups, *t* test: $t_{(68)} = 0.70$, $p(>.10)$ and Mann–Whitney *U*: $U = 356.00$, $Z = -0.389$, $p(>.10)$.

6 | DISCUSSION

The aim of this study was to investigate the role of a construct-based SJT as a preselection tool for the outcome of an AC in the context of a multistage selection process for pilots. However, low positive relationships with knowledge tests and no significant correlations with cognitive performance tests or personality scales were found (see Table 1). As the SJT-TW was carefully developed to measure teamwork behavior (Gatzka & Volmer, 2017), its validity does not come into question due to these low correlations. On the contrary, the latter can be interpreted as reflecting the SJT-TW's capability to cover the aspects of candidates' characteristics that differ from the constructs assessed thus far. Furthermore, as knowledge-based instruction was used, the higher correlations with knowledge tests corresponded with earlier findings (e.g., Lievens & Patterson, 2011).

As expected, significant positive correlations were found between the SJT-TW and behavioral observations from the AC, especially for leadership, teamwork/communication, and

TABLE 3 Results of the stepwise multiple regression for the prediction of dimensionwise assessment center performance

	R	R^2 (R^2 corr.)	ΔR^2	F	p
Leadership					
Model 1	.25	.06 (.04)	.06**	3.06	.006
Model 2	.31	.10 (.06)	.03	2.80	.003
Model 3	.39	.16 (.12)	.06***	4.42	<.001
Teamwork/communication					
Model 1	.29	.08 (.06)	.08**	4.00	.001
Model 2	.30	.09 (.05)	.01	2.57	.006
Model 3	.37	.13 (.10)	.04***	3.68	<.001
Adherence to procedures					
Model 1	.21	.04 (.02)	.04	1.98	.068
Model 2	.26	.07 (.03)	.03	1.96	.038
Model 3	.27	.07 (.03)	.00	1.83	.050
Resilience					
Model 1	.22	.05 (.03)	.05*	2.30	.035
Model 2	.26	.07 (.03)	.02	1.98	.036
Model 3	.37	.14 (.10)	.07***	3.79	<.001

Note: Model 1 includes the predictor variables English, technics, mathematics, memory, concentration and spatial orientation; Model 2 additionally includes extraversion, conscientiousness, agreeableness and neuroticism; Model 3 additionally includes the situational test for teamwork.

*p < .05; **p < .01; ***p < .001.

resilience (see Table 2). Initially, these correlations between the SJT-TW and leadership and resilience might be striking, especially since they surpass the correlation between the SJT-TW and teamwork/communication. However, considering the design background of both instruments, this effect can be explained as follows: While the AC was developed to discriminate between leadership and teamwork/communication (for pilot selection), the SJT-TW was developed independently of the AC and its application context, incorporating leadership facets as part of the construct "teamwork." As Gatzka and Volmer (2017) reported, significant correlations were found for $N = 118$ employees between the SJT-TW and leadership ($r = .24$), coordination ($r = .25$) and decision making ($r = .25$). These facets of teamwork are explicit parts of the dimension of leadership as conceptualized in our AC (see above). Resilience was also significantly correlated with the SJT (and with all AC dimensions; see Table 1). In addition, for the applicants in our study, the stakes were high, so there existed the potential for higher degrees of stress. As was confirmed earlier, assessors can observe stress symptoms correctly, and these symptoms can impair AC performance (McClimon, 2018). Therefore, we concluded that the SJT-TW has the potential to assess teamwork (and leadership), which are relevant parts of applicants' "real" behavior.

Furthermore, the explanatory power of the SJT-TW for the probability of the outcome of the AC (Table 3) and single AC dimensions (Table 4) was confirmed. The contribution of the SJT-TW was incremental compared to the tests administered in the first selection stage and these results matched the effects that McDaniel et al. (2007) found in their meta-analysis concerning the incremental validity of SJTs over personality and cognitive ability. However, notably, our "criterion" was only cross-sectional and not a job performance measure.

The variables "memory" and "agreeableness" proved to be relevant for passing the AC, which might be explained by its requirements: For successful performance in the single tasks, it was important to memorize complex instructions and rules while taking part agreeably in the social interaction. However, extraversion—which is frequently found to be correlated with AC outcomes (Collins et al., 2003)—was not relevant in the regression equation, but we found a significant bivariate correlation with the AC dimension of leadership ($r = .17$).

Overall, our results are in line with those of Lievens and Patterson (2011). We expanded and specified their findings by confirming the relationship between an AC and an SJT on a construct level following the request of several researchers for a construct-based analysis of selection instruments (e.g., Arthur & Villado, 2008; McDaniel et al., 2007; Ployhart, 2006; Roth et al., 2008). Our results also contribute to more transparency, as they allow for a detailed interpretation of the relationship between the SJT-TW and teamwork-relevant AC dimensions.

In our study, we used an SJT that was developed for the general measurement of teamwork, regardless of specific job profiles. By utilizing an SJT constructed in parallel to our AC dimensions, even higher correlations can be expected (see also Lievens & Patterson, 2011). Furthermore, theoretically, our results confirmed the usefulness of low-fidelity simulations for the assessment of behavioral constructs. Practically, our findings also supported the idea of purposefully using construct-based SJTs as preselection tools. This is in line with the approach of basing personnel selection on relevant constructs (e.g., regarding job profiles or developmental purposes; Caughron et al., 2012). For applications in the personnel selection process, there are several advantages of using a construct-based SJT as a preselection tool for an AC: First, SJTs are more efficient in terms of the number of applicants who can be tested simultaneously and at different locations (online). Second, observers are not involved, making it more efficient and less costly while minimizing the risk of observer biases. Third, SJTs allow for greater flexibility in simulating different job-relevant situations: Aside from the assessment of a higher number of different job-relevant situations, it is possible to assess situations that cannot be operationalized in an AC due to ethical or other practical reasons.

While interpreting the results presented here, it has to be considered that the setting was a high-stakes selection context. The sample was, thus, selected to meet predefined requirements (e.g., all candidates had to have completed high-school education that met university entrance requirements; their age was younger than

30 years; they had to have passed the first stage of pilot selection). This led to restricted variance in the data set and might have obscured stronger effects. Finally, to confirm the usefulness of administering SJTs before an AC, it is necessary to determine their criterion-oriented validity for later training or job performance. However, in this study, only initial tentative hints for the predictive validity of the SJT-TW could be reported.

ACKNOWLEDGEMENT

Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

Research data are not shared.

ORCID

Panja Goerke http://orcid.org/0000-0002-0340-388X
Julia Maier http://orcid.org/0000-0003-1387-7939

REFERENCES

Arthur, W., Jr., & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology*, *93*(2), 435–442. https://doi.org/10.1037/0021-9010.93.2.435

Arthur, W., Jr., 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*, 125–154. https://doi.org/10.1111/j.1744-6570.2003.tb00146.x

Caughron, J. J., Mumford, M. D., & Fleishman, E. A. (2012). The Fleishman Job Analysis survey: Development, validation, and applications. In M. A. Wilson, W. Bennett, Jr., S. G. Gibson, & G. M. Alliger (Eds.), *The handbook of work analysis: Methods, systems, applications and science of work measurement in organizations* (pp. 231–246). Routledge: Taylor & Francis Group.

Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology*, *63*(1), 83–117. https://doi.org/10.1111/j.1744-6570.2009.01163.x

Collins, J. M., Schmidt, F. L., Sanchez-Ku, M., Thomas, L., McDaniel, M. A., & Le, H. (2003). Can basic individual differences shed light on the construct meaning of assessment center evaluations. *International Journal of Selection and Assessment*, *11*(1), 17–29. https://doi.org/10.1111/1468-2389.00223

Gatzka, T., & Volmer, J. (2017). Situational judgment test für teamarbeit (SJT-TA) [Situational judgment test for teamwork (SJT-TW)]. *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*. https://doi.org/10.6102/zis249

Goeters, K. M., Timmermann, B., & Maschke, P. (1993). The construction of personality questionnaires for selection of aviation personnel. *International Journal of Aviation Psychology*, *3*(2), 123–141. https://doi.org/10.1207/s15327108ijap0302_3

Hair, J. F., Babin, B. J., Anderson, R. E., & Black, W. C. (2018). *Multivariate data analysis*. Cengage Learning, EMEA.

Kasten, N., Freund, P. A., & Staufenbiel, T. (2018). "Sweet lies": An in-depth analysis of faking behavior on situational judgment tests compared to personality questionnaires. *European Journal of Psychological Assessment*, *36*(1), 136–148. https://doi.org/10.1027/1015-5759/a000479

Lievens, F., & De Soete, B. (2015). Situational judgment test. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (Vol. 22, pp. 13–19). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.25092-7

Lievens, F., & Patterson, F. (2011). The validity and incremental validity of knowledge tests, low-fidelity simulations, and high-fidelity simulations for predicting job performance in advanced-level high-stakes selection. *Journal of Applied Psychology*, *96*(5), 927–940. https://doi.org/10.1037/a0023496

Lievens, F., Schäpers, P., & Herde, C. N. (2020). Situational Judgment Tests: From low-fidelity simulations to alternative measures of personality and the person–situation interplay. In D. Wood, P. Harms, S. Read, & A. Slaughter (Eds.), *Emerging approaches to measuring and modeling the person and situation* (pp. 285–311). Elsevier. https://doi.org/10.1016/B978-0-12-819200-9.00017-X

McClimon, M. (2018). *The effect of anxiety on assessment center performance* (Masters thesis) ProQuest Dissertations Publishing. https://www.proquest.com/dissertations-theses/effect-anxiety-on-assessment-center-performance/docview/2051869610/se-2

McDaniel, M. A., & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment*, *9*(1–2), 103–113. https://doi.org/10.1111/1468-2389.00167

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, III W. L. (2007). Situational judgment tests, response instruction, and validity: A meta-analysis. *Personnel Psychology*, *60*(1), 63–91. https://doi.org/10.1111/j.1744-6570.2007.00065.x

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology*, *86*(4), 730–740. https://doi.org/10.1037/0021-9010.86.4.730

Mittelstädt, J. M., Pecena, Y., Oubaid, V., & Maschke, P. (2016). Construct validity of the temperament structure scales within the Big Five framework in aerospace selection. *Aviation Psychology and Applied Human Factors*, *6*(2), 68–80. https://doi.org/10.1027/2192-0923/a000101

Nguyen, N. T., Biedermann, M. D., & McDaniel, M. A. (2005). Effects of response instructions on faking a situational judgment test. *International Journal of Selection and Assessment*, *13*(4), 250–260. https://doi.org/10.1111/j.1468-2389.2005.00322.x

Ployhart, R. (2006). Staffing in the 21st century: New challenges and strategic opportunities. *Journal of Management*, *32*(6), 868–897. https://doi.org/10.1177/0149206306293625

Roth, P., Bobko, P., McFarland, L., & Buster, M. (2008). Work sample tests in personnel selection: A meta-analysis of Black–White differences in overall and exercise scores. *Personnel Psychology*, *61*, 637–661. https://doi.org/10.1111/j.1744-6570.2008.00125.x

Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2021). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology*. Advance online publication. https://doi.org/10.1037/apl0000994

Thornton, G. C. III, & Rupp, D. (2006). *Assessment centers in human resource management: Strategies for prediction, diagnosis, and development*. Erlbaum.

Wheekley, J. A., Hawkes, B., Guenole, N., & Ployhart, R. E. (2015). Low-fidelity simulations. *Annual Review of Organizational Psychology and Organizational Behavior*, *2*, 295–322. https://doi.org/10.1146/annurev-orgpsych-032414-111304

Zhang, C., Cullen, M. J., & Sackett, P. R. (2021). Effects of response instructions on situational judgment test performance in operational selection and developmental contexts. *International Journal of Selection and Assessment*, *29*(2), 170–184. https://doi.org/10.1111/ijsa.12324

Zinn, F., Goerke, P., & Marggraf-Micheel, C. (2020). Selecting for cockpit crew. In R. Bor, C. Eriksen, T. L. Hubbard, & R. King (Eds.), *Pilot selection—Psychological principles and practice* (pp. 21–34). CRC Press, Taylor & Francis Group.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Goerke, P., & Maier, J. (2022). Analysis of a situational judgment test for teamwork as a preselection tool for an assessment center: A construct-based approach. *International Journal of Selection and Assessment*, 30, 456–464. https://doi.org/10.1111/ijsa.12391