

Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection Lee Kong Chian School Of
Business

Lee Kong Chian School of Business

8-2020

Effects of situation descriptions on the construct-related validity of construct-driven situational judgment tests

Philipp Christopher SCHAEPPERS

Singapore Management University, pschappers@smu.edu.sg

Jan-Philipp FREUDENSTEIN

Freie Universitat Berlin

Patrick MUSSEL

Freie Universitat Berlin

Filip LIEVENS

Singapore Management University, filiplievens@smu.edu.sg

Stefan KRUMM

Freie Universitat Berlin

Follow this and additional works at: https://ink.library.smu.edu.sg/lkcsb_research



Part of the [Industrial and Organizational Psychology Commons](#), and the [Organizational Behavior and Theory Commons](#)

Citation

SCHAEPPERS, Philipp Christopher; FREUDENSTEIN, Jan-Philipp; MUSSEL, Patrick; LIEVENS, Filip; and KRUMM, Stefan. Effects of situation descriptions on the construct-related validity of construct-driven situational judgment tests. (2020). *Journal of Research in Personality*. 87, 1-5.

Available at: https://ink.library.smu.edu.sg/lkcsb_research/6600

This Journal Article is brought to you for free and open access by the Lee Kong Chian School of Business at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection Lee Kong Chian School Of Business by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylds@smu.edu.sg.

Effects of situation descriptions on the construct-related validity of construct-driven situational judgment tests

Philipp Schäpers^{a1}, Jan-Philipp Freudenstein^{b1}, Patrick Mussel^b Filip Lievens^a Stefan Krumm^b

^a Lee Kong Chian School of Business, Singapore Management University, Singapore

^b Institute of Psychology, Freie Universität Berlin, Germany

Published in *Journal of Research in Personality*, 2020 August, Volume 87, Article Number 103963.

<https://doi.org/10.1016/j.jrp.2020.103963>

Abstract

Despite the common belief that situation descriptions in SJTs are central to the measurement of targeted constructs, recent studies demonstrated that omitting situation descriptions had only minor effects on SJT performance and validity. However, these results might be due to the fact that traditional SJTs often fail to assess well-defined constructs. So, we aimed to examine the relationships between construct-driven SJT scores with and without situations and self- and peer-rated personality dimensions (N = 158). Results revealed almost no difference in construct-related validity between both versions. The conscientiousness facet emerged as the only exception, for which the SJT scores without situation descriptions yielded a significantly higher convergent correlation.

Keywords

Situational judgment test, Construct-related validity, Contextualization, Personality assessment

1. Introduction

In the last decades, there has been growing interest in developing Situational Judgment Tests (SJTs), because they come with several favorable properties (e.g., they are more difficult to fake than self-reports, Kasten, Freund, & Staufenbiel, 2018). Test-takers' task is to indicate how they would/ should react to a situation. Thus, situation descriptions in SJTs are believed to play an important role and many SJT researchers accordingly described them as core element of any SJT (e.g., Campion & Ployhart, 2013).

However, recent studies (e.g., Krumm et al., 2015, Schäpers, Mussel et al., 2019) put a crack in this edifice of SJTs. Specifically, Krumm et al. discovered that a considerable number of SJT items (between 43 and 71%) could be solved correctly regardless of the availability of situation descriptions in the item stems.¹ A recent study substantiated these results and found no or only small differences in construct-related validity for three traditional (work-related) SJTs when applied with or without situation descriptions (Schäpers, Mussel et al., 2019). However, these traditional SJTs are often described as multidimensional measures that fail to assess single constructs. Consequently, their construct-related validity was recently described as a "hot mess" (McDaniel, List, & Kepes, 2016, p. 47). Thus, one may argue that traditional SJTs with and without situation descriptions did not show differences in construct-related validity because these SJTs did not provide satisfactory construct-related validity in the first place.

¹ Philipp Schäpers and Jan-Philipp Freudenstein contributed equally to this work. All data and code are available on the Open Science Framework (<https://osf.io/py6cm/>).

Thus, several scholars have called for a more construct-driven approach when developing SJTs (e.g., Guenole et al., 2017, Lievens, 2017). In this line of research, SJTs are developed as measures of personality constructs. This led to SJTs that show a satisfactory factor structure with a clear pattern of convergent and discriminant validity (e.g., Mussel, Gatzka, & Hewig, 2018, for an example item of a construct-driven SJT see Electronic Supplementary Material Fig. S1). Building on Trait Activation Theory, situation descriptions in construct-driven SJTs contain relevant cues that activate trait-relevant behavior (Mussel et al., 2018). It is important that the trait-related situational cues are of moderate strength; that is, they should not activate the same trait-related behavior in every test-taker (Harris, Siedor, Fan, Listyg, & Carter, 2016). In the same vein, Marshall and Brown (2006) found that for people who are high on a trait, only a moderate situational strength is needed to arouse a trait-relevant response. Thus, strong situations (e.g., a red traffic light) leave too little of room for variability in trait-relevant behavior, whereas moderate and “weak” situations (e.g., yellow traffic light) leave room to observe variability. Initial evidence revealed that this rationale applies to construct-driven SJTs (Schäpers, Lievens et al., 2019). These authors showed that the inclusion of trait-related situational cues affects trait-related responses in SJTs.

This study examined whether the presence of situation descriptions in the item stems affects the construct-related validity of construct-driven SJT scores. Against the backdrop of the above, situation descriptions should be especially relevant in construct-driven SJTs. Thus:

H1: The construct-related validity of scores on construct-driven SJTs will be significantly higher for SJTs with situation descriptions than for SJTs without situation descriptions.

The current study extends previous studies on the relevance of situation descriptions in several ways. First, we focus on a construct-driven SJT that gauges facets of personality. Second, the herein applied SJT showed strong evidence of convergent and divergent validity. In the current study, we can therefore expect to find similar patterns; and mismatching patterns between the version with vs. without situation descriptions will also be insightful (and not just evidence of the typical “hot mess” in SJTs; McDaniel et al., 2016, p. 47). Third, no previous study on the relevance of situation descriptions in SJTs has used a within-subjects design. Thus, past research could not rule out that (lack of) differences between SJTs with vs. without situation descriptions were driven by individual difference factors that affected SJT response behavior.

2. Method

2.1. Participants

We expected the SJT with situation descriptions to have a construct-related validity of $r = 0.50$ (see Mussel et al., 2018). Situation descriptions in construct-driven SJTs serve a trait-activating purpose and are thus typically regarded as essential elements for the construct-related validity of the SJT scores. Hence, we determined differences of $r = 0.20$ between the condition with situation descriptions and the condition without situation descriptions with NEO-PI-R self-reports as meaningful. An a-priori power-analysis with G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) revealed that $N = 152$ participants are necessary to detect such a within-subject differences of correlation coefficients ($\alpha = 0.05$; $1 - \beta = 0.80$; two-tailed z-test for differences between correlations of two dependent correlations with common index and an assumed correlation of $r = 0.50$ between conditions). The sample comprised 158 individuals (59.5% female). Participants were on average 36.86 years old ($SD = 14.72$, range 19 to 71) and were recruited by students undertaking an undergraduate class at a German comprehensive state university. We also asked the undergraduate students to serve as peer raters. Peer-rated data were available for all except one person. Peers (83.3% female) were on average 23.29 years old ($SD = 2.76$, range 20 to 31) and knew the participant on average for 14.22 years ($SD = 9.30$).

2.2. Study design and materials

All data were collected online. There were two parallel versions (A and B) of the same SJT. We adopted a within-subjects design: All participants worked on an SJT version with situation descriptions *and* on a

(parallel) SJT version without situation descriptions (for an overview of the test procedure see Electronic Supplementary Material Fig. S2).

Situational judgement test. The Big Five Situational Judgment Test (Mussel et al., 2018) served as SJT in our study. It assesses five lower-order personality facets, namely gregariousness (extraversion), compliance (agreeableness), self-discipline (conscientiousness), self-consciousness (neuroticism), and openness to ideas (openness to experience), with 22 items per facet. Each item consists of a situation description tapping into one of the personality facets along, and four response options. When working on the SJT version without situation descriptions, we asked participants only “What would you do?” and presented response options without any situation content (see Electronic Supplementary Material Fig. S1). For each item, two of the response options reflected behavior typical for individuals with either high or low levels of a particular trait. Participants’ responses were either scored as “1” or “0”, depending on whether they chose response options reflecting high or low trait expressions.

According to prior research, this SJT exhibits convincing convergent ($r_s = 0.41 - 0.70$) and discriminant validity ($r_s = -0.35 - 0.19$) when correlated with traditional personality inventories assessing these five facets (Mussel et al., 2018). On the basis of a reanalysis of Mussel et al.’s data, we created two parallel SJT versions with 11 items per facet. This was done by identifying pairs of similar items based on item difficulties, standard deviations, and corrected item-scale correlations, which we assigned to the two SJT versions. Reliability estimates of both SJT versions ranged from $\omega = 0.34$ to 0.75 (see Table 1).² For details on the analytical strategy, a test of parallelism, and a test of metric measurement invariance, see Electronic Supplementary Material (p. 1–3).

Table 1. Descriptive statistics and reliability estimates of the big five SJT.

Facet	SJT with situation descriptions (n = 158)					SJT without situation descriptions (n = 158)				
	M	SD	ω	skewness	kurtosis	M	SD	ω	skewness	kurtosis
Gregariousness (extraversion)	0.38	0.20	0.63 [0.52, 0.74]	0.55	-0.21	0.43	0.22	0.66 [0.57, 0.74]	0.25	-0.72
Compliance (agreeableness)	0.60	0.16	0.43 [0.19, 0.67]	0.32	-0.26	0.61	0.18	0.51 [0.37, 0.65]	-0.38	-0.16
Self-discipline (conscientiousness)	0.56	0.19	0.57 [0.46, 0.69]	-0.33	-0.20	0.54	0.20	0.54 [0.42, 0.65]	-0.18	-0.69
Self-consciousness (neuroticism)	0.29	0.16	0.37 [0.12, 0.69]	0.41	-0.04	0.36	0.19	0.58 [0.47, 0.70]	0.35	-0.40
Openness for ideas (openness)	0.56	0.18	0.57 [0.45, 0.69]	-0.19	-0.23	0.60	0.18	0.55 [0.43, 0.66]	-0.10	-0.58

Note. ω = Omega total; 95% confidence intervals in brackets.

Big Five personality dimensions. The same lower-order personality facets as in the SJT were assessed with the German version of the NEO-PI-R personality questionnaire (Ostendorf & Angleitner, 2004). Each facet consisted of eight items. Participants responded on a 7-point Likert scale ranging from 1 = *disagree strongly* to 7 = *agree strongly*. Reliability for the facet scores ranged from $\omega = 0.59$ to 0.85. The same scales were used to collect peer-rated personality (ω ranged from 0.74 to 0.90; see Table S1).

3. Results

Similar to Schäpers, Mussel et al., 2019, we started by comparing mean scores for the SJT with vs. without situation descriptions. Availability of the situation descriptions either did not make a difference, which was the case for compliance, $t(1\ 5\ 7) = -0.63$, $p = .53$, $d = 0.05$, and self-discipline, $t(1\ 5\ 7) = 0.08$, $p = .30$, $d = -0.08$, or even led to lower scores of SJTs with situation descriptions as compared to SJTs without situation descriptions, as was the case for gregariousness, $t(1\ 5\ 7) = -3.59$, $p < .001$, $d = -0.29$, self-consciousness, $t(1\ 5\ 7) = -5.24$, $p < .001$, $d = -0.42$, and openness to ideas, $t(1\ 5\ 7) = -3.21$, $p < .01$, $d = -0.26$. At the item level, it did not make a significant difference whether the situation description was presented or not for between 86 and 96% of the items—depending on the correction for alpha inflation (we deliberately report the range from no correction to Bonferroni correction; for details, see Electronic Supplementary Material Table S3).

The mean convergent correlation between the SJT with situation descriptions and the personality self-ratings was $r = 0.45$. Notably, almost the same mean correlation ($r = 0.46$) was found for the SJT without situation descriptions ($z = -0.14, p = .88$). Comparisons at the facet level revealed that, for four out of five facets, the convergent validity correlation coefficients with self-rated personality did not differ across both SJT versions (z s = -0.72 to $1.26, p$ s = 0.207 to 0.796 ; see Table 2). The conscientiousness facet (self-discipline) was the only exception, for which the SJT without situation descriptions yielded a significantly higher convergent correlation ($r = 0.53$) than the SJT with situation description ($r = 0.34, z = 2.79, p = .005$), thereby lending no support for H1.

Table 2. Correlations of the Big Five SJT (With and Without Situation Descriptions) with NEO-PI-R Self-reports and Peer Ratings.

Facet	SJT with situation description					SJT without situation description				
	1	2	3	4	5	6	7	8	9	10
<i>Self-reported personality</i>										
1. Gregariousness	0.60**	0.05	0.04	-0.19*	0.11	0.64**	0.11	0.04	-0.15	0.19*
2. Compliance	-0.13	0.29**	0.15	-0.03	0.08	0.02	0.17*	0.09	-0.08	0.09
3. Self-discipline	-0.01	0.10	0.34**	-0.26**	0.12	-0.01	-0.03	0.53**	-0.11	0.15
4. Self-consciousness	-0.17*	0.15	-0.08	-0.47**	-0.16	-0.22**	0.02	-0.19*	0.38**	-0.16*
5. Openness for ideas	0.20*	0.01	0.03	-0.30**	0.53**	0.08	0.02	-0.10	-0.24**	0.52**
<i>Peer-rated personality</i>										
6. Gregariousness	0.29**	-0.08	0.08	-0.17*	0.10	0.42**	0.04	0.05	-0.13	0.09
7. Compliance	-0.11	0.08	0.00	-0.04	0.04	0.06	0.27**	0.03	0.01	0.04
8. Self-discipline	-0.09	-0.02	0.23**	-0.03	-0.12	-0.11	-0.03	0.24**	0.04	-0.07
9. Self-consciousness	-0.15	0.26**	0.00	0.28**	-0.02	-0.10	0.07	-0.03	0.30**	-0.10
10. Openness for ideas	0.04	-0.14	-0.05	-0.17*	0.14	0.02	0.02	-0.05	-0.20**	0.24**

Note. $N = 157-158$. * $p < .05$. ** $p < .01$.

Analyses for peer-rated personality yielded similar results. The mean convergent validity between the SJT and peer-rated personality was $r = 0.21$ when situation descriptions were presented and $r = 0.29$ when situation descriptions were omitted ($z = -1.01, p = .314$). At the facet level, the SJT without situation descriptions showed no significant difference in convergent validity coefficients as compared to the SJT with situation descriptions.

Finally, to further scrutinize differences in construct-related validity between SJT versions with and without situation descriptions, we also ran a CTC(M-1) model (Eid, Lischetzke, Nussbeck, & Trierweiler, 2003). Manifest facet scores of the NEO-PI-R (self-reported) and both SJT versions (with and without situation descriptions) were used as indicators for the corresponding traits. Additionally, we specified two latent method factors for SJT scores with and without situation description. The NEO-PI-R facets were set as reference factor (Fig. 1; see Eid et al., 2003). This model showed good fit, $\chi^2(69) = 94.66, p = .022$, RMSEA = 0.049, SRMR = 0.061, CFI = 0.955. All loadings on the latent facet factors were significant with the exception of the SJT compliance score without situation description. On average, loadings of SJT scores with situation descriptions on the latent facet factors were $\lambda = 0.70$, whereas loadings of SJT scores without situation descriptions were on average $\lambda = 0.66$. The biggest difference emerged between the loadings of the SJT scores for compliance ($\lambda_{\text{with Situation}} = 0.88; \lambda_{\text{without Situation}} = 0.24$). Yet, when restraining the factor loadings of SJT facets with and without situation descriptions to equality, the model fit did not decrease significantly, $\Delta\chi^2(5) = 9.80, p = .081, \Delta\text{RMSEA} = 0.001, \Delta\text{SRMR} = 0.003, \Delta\text{CFI} = 0.006$, as compared to the unconstrained model. Similar results were obtained when peer-ratings were used as additional indicators, $\Delta\chi^2(5) = 5.73, p = .33, \Delta\text{RMSEA} = 0.003, \Delta\text{SRMR} = 0.001, \Delta\text{CFI} = 0.010$. Yet, this model did not reveal a satisfactory model fit (RMSEA = 0.071, SRMR = 0.068, CFI = 0.850).

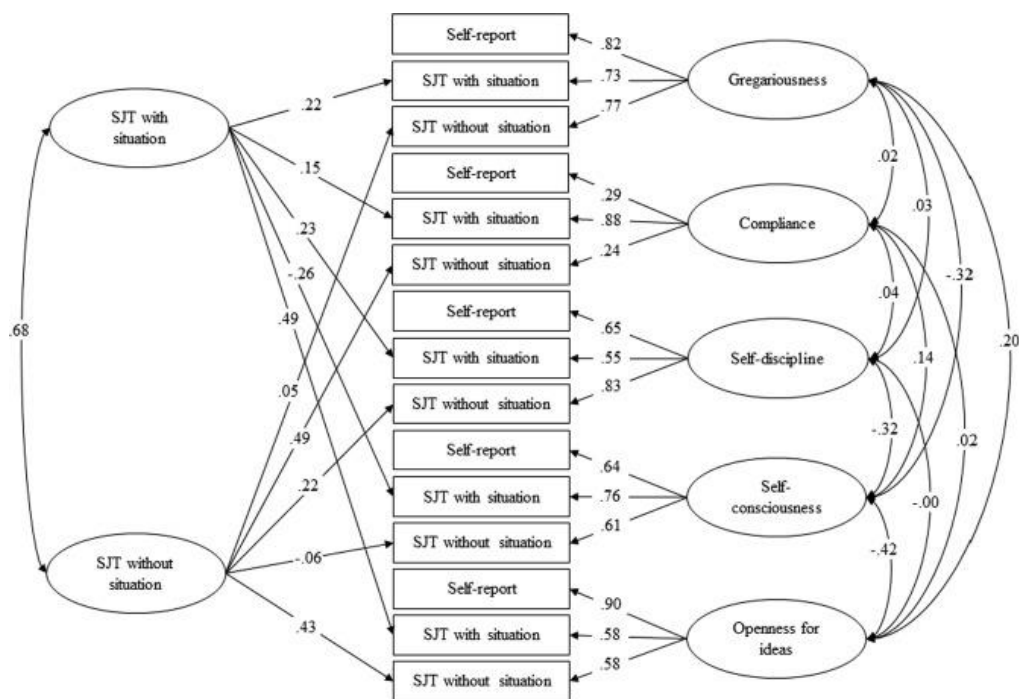


Fig. 1. CTC(M-1) model with manifest facet scores of the NEO-PI-R (self-reported) and both SJT versions (with and without situation descriptions) as indicators for the corresponding trait-facets. Two latent method factors were specified for SJT scores with and without situation descriptions. The NEO-PI-R facets were set as reference factor; $\chi^2(69) = 94.66$, $p = .022$, RMSEA = 0.049, SRMR = 0.061, CFI = 0.955.

4. Discussion

The current study examined whether the presence of situation descriptions affected the construct-related validity of construct-driven SJT scores. Notably, our study extended prior research by testing differences in construct-related validity for a construct-driven SJT (cf. Schäpers, Mussel et al., 2019). We found little support for differences in construct-related validity among SJT scores with and without situation descriptions in the item stems. This was true (i) for mean convergent validity coefficients, (ii) for coefficients inspected at the facet level as well as (iii) across raters (self- versus peer-ratings). Thus, the use of SJTs to assess interactional processes of personality and situations may be questioned and open to further examination.

The SJT score for the conscientiousness facet of self-discipline formed the only exception: There was a significantly higher correlation with self-reported self-discipline when the SJT administered without situation descriptions. The direction of this effect is contrary to our hypothesis. One explanation is that situation descriptions included in the SJT may have activated not only trait-related behavior, but also construct-irrelevant behavior. This effect may have been attenuated in the SJT version without situation descriptions. In other words, SJT items without situation description may be more similar to classical self-report items, which may enhance convergent correlations. However, as the increased convergent correlation was evident for only one facet and only for self-rated and not for peer-rated data, this explanation should be interpreted with caution. More research is needed to assess whether peer-raters rely on situation descriptions of SJTs when rating target individuals.

Our findings have several implications for SJT research and theory. Previous research in this field relied on between-subject designs (e.g., Schäpers, Mussel et al., 2019). This study is the first to contrast SJTs with vs. without situation descriptions in a within-subject design. Therefore, it could be ruled out that (lack of) differences between SJTs with vs. without situation descriptions were driven by individual difference factors that affected SJT response behavior. That being said, our results also showed that the overlap between SJT versions with and without situation descriptions is far from being perfect (see Electronic Supplementary Material Table S2). Although the construct-related validity did mostly not differ, unique portions of variance

for each test version may contain different and valuable information. Future research needs to link these variance components to real-life behaviors, external criteria and knowledge components that have been deemed important in the context of traditional SJTs. For instance, in a recent study, Schäpers et al. demonstrated that SJT versions with and without situation descriptions did not differ in predicting global job performance criteria (e.g., OCBI), but they did differ in predicting specific job performance criteria (e.g., interpersonal adaptability).

Finally, this study provides further evidence that experimental test validation provides interesting insights that go beyond correlative validation approaches. Thus, we also add to a growing body of experimental validation research, which is not limited to SJTs (for a further example in the domain of reading comprehension see Schroeder & Tiffin-Richards, 2014).

In short, our study adds to the current debate about the conceptualization of SJT as context-(in)dependent measures. Despite the findings of Krumm et al. (2015), several researchers posited that SJTs measure situational judgment (e.g., McDaniel et al., 2016, Harris et al., 2016). Specifically, Harris et al. posited that situational judgment can be applied only if individuals use situational cues for effective trait expression. However, by showing that the construct-related validity of SJTs is generally not contingent on situation descriptions being presented or not, this study adds more evidence speaking to the perspective of SJTs as context-independent measures (see also Lievens & Motowidlo, 2016).

5. Limitations

As a first limitation, we used only one SJT from one construct domain. So, future research is needed to examine the generalizability of our findings to other construct-driven SJTs. Second, due to the relatively small sample size, we were not able to assess the fit of the SJT measurement models. Such tests could have strengthened interpretations with regard to mean comparisons between SJT facets with and without situation descriptions. Given the varying internal consistencies of SJT scores in general, future research needs to address the question of adequate measurement models for SJTs. Finally, the herein used forced-choice response format comes with both advantages and drawbacks compared to other SJT scoring formats (see Olaru, Jankowsky, Mussel, Mazziotta, 2019). Regarding correlations with other measures, this format leads to variance restriction which might reduce correlations with other methods. So, we recommend replicating this study with SJTs based on other scoring keys.

6. Conclusion

Our study's key message is that the absence of SJT situation descriptions did not or only marginally affect the construct-related validity of construct-driven SJT scores. We encourage researchers to study possible moderators that may explain effects of SJT situations on SJT scores' construct-related validity.

References

- Campion, M. C., & Ployhart, R. E. (2013). Assessing personality with situational judgment measures: Interactionist psychology operationalized. In N. D. Christiansen & R. P. Tett (Eds.), *Handbook of personality at work* (pp. 439–456). New York, NY: Routledge. <https://doi.org/10.4324/9780203526910.ch19> .
- Eid, M., Lischetzke, T., Nussbeck, F. W., & Trierweiler, L. I. (2003). Separating trait effects from trait-specific method effects in multitrait-multimethod models: a multiple-indicator CT-C(M-1) model. *Psychological Methods*, 8, 38–60. <https://doi.org/10.1037/1082-989x.8.1.38> .
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191. <https://doi.org/10.3758/BF03193146> .

- Guenole, N., Chernyshenko, O. S., & Weekly, J. (2017). On designing construct driven situational judgment tests: Some preliminary recommendations. *International Journal of Testing*, 17, 234–252. <https://doi.org/10.1080/15305058.2017.1297817> .
- Harris, A. M., Siedor, L. E., Fan, Y., Listyg, B., & Carter, N. T. (2016). In defense of the situation: An interactionist explanation for performance on situational judgment tests. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 9, 23–28. <https://doi.org/10.1017/iop.2015.110>.
- Kasten, N., & Freund, P. A. (2016). A meta-analytical multilevel reliability generalization of situational judgment tests (SJTs). *European Journal of Psychological Assessment*, 32, 230–240. <https://doi.org/10.1027/1015-5759/a000250>.
- Kasten, N., Freund, P. A., & Staufienbiel, T. (2018). Sweet little lies. An in-depth analysis of faking behavior on situational judgment tests compared to personality questionnaires. *European Journal of Psychological Assessment*. <https://doi.org/10.1027/1015-5759/a000479>.
- Krumm, S., Lievens, F., Hüffmeier, J., Lipnevich, A. A., Bendels, H., & Hertel, G. (2015). How “situational” is judgment in situational judgment tests?. *Journal of Applied Psychology*, 100, 399–416. <https://doi.org/10.1037/a0037674> .
- Lievens, F. (2017). Construct-driven SJTs: Toward an agenda for future research. *International Journal of Testing*, 17, 269–276. <https://doi.org/10.1080/15305058.2017.1309857>.
- Lievens, F., & Motowidlo, S. J. (2016). Situational judgment tests: From measures of situational judgment to measures of general domain knowledge. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 9, 3–22. <https://doi.org/10.1017/iop.2015.71> .
- Marshall, M. A., & Brown, J. D. (2006). Trait aggressiveness and situational provocation: A test of the traits as situational sensitivities (TASS) model. *Personality and Social Psychology Bulletin*, 32, 1100–1113. <https://doi.org/10.1177/0146167206288488> .
- McDaniel, M. A., List, S. K., & Kepes, S. (2016). The “hot mess” of situational judgment test construct validity and other issues. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 9, 47–51. <https://doi.org/10.1017/iop.2015.115> .
- Mussel, P., Gatzka, T., & Hewig, J. (2018). Situational judgment tests as an alternative measure for personality assessment. *European Journal of Psychological Assessment*, 34, 328–335. <https://doi.org/10.1027/1015-5759/a000346> .
- Ostendorf, F., & Angleitner, A. (2004). NEO Personality Inventory based on Costa and McCrae, revised version (NEO-PI-R). Göttingen, Germany: Hogrefe.
- Schäpers, P., Lievens, F., Freudenstein, J.-P., Schulze, J., König, C. J., & Krumm, S. (2019). Which Kind of Situational Information is Needed to Make Situational Judgment Tests Situational? Presented at the 19th convention of the European Association of Work and Organizational Psychology (EAWOP), Turin, Italy, 29 May – 1 June 2019. doi:10.13140/RG.2.2.32935.50081.
- Schäpers, P., Mussel, P., Lievens, F., König, C. J., Freudenstein, J.-P., & Krumm, S. (2019). The role of situations in situational judgment tests: Effects on construct saturation, predictive validity, and applicant perceptions. *Journal of Applied Psychology*. <https://doi.org/10.1037/apl0000457> .
- Schroeder, S., & Tiffin-Richards, S. (2014). Cognitive processing of reading comprehension items with and without text. *Zeitschrift Für Pädagogische Psychologie*, 28, 21–30. <https://doi.org/10.1024/1010-0652/a00121> .