

Electronic Supplementary Material for

Ohlms, M. L., Voigtländer, E., Melchers, K. G., & Kanning, U. P. (2023). Is gamification a suitable means to improve applicant reactions and convey information during an online test? *Journal of Personnel Psychology*. <https://doi.org/10.1027/1866-5888/a000343>

Table E1 shows all items used to measure the different applicant reaction variables.

Confirmatory Factor Analysis Related to the Applicant Reaction Variables

We conducted a confirmatory factor analysis (CFA) with seven correlated factors, to test whether the different perception variables from the post questionnaire indeed represented separable constructs. To do so, we used the *lavaan* package in R (Version 4.3.1, Rosseel, 2012). This CFA revealed adequate fit, $\chi^2(209) = 357.93, p < .001$, CFI = .95, SRMR = .05, RMSEA = .06. In contrast, a single-factor model fitted poorly, $\chi^2(230) = 1962.79, p < .001$, CFI = .40, SRMR = .14, RMSEA = .19. Additionally, the seven-factor model showed a significantly better fit than the single-factor model, $\Delta\chi^2(21) = 1604.86, p < .001$. Similarly, for the two reaction variables that were assessed prior to administering the cognitive ability test, we tested a CFA model with either two correlated factors or a single-factor model. The two-factor model fitted well, $\chi^2(8) = 12.73, p = .12$, CFI = .99, SRMR = .04, RMSEA = .05, whereas the single-factor model had a poor fit, $\chi^2(9) = 133.98, p < .001$, CFI = .76, SRMR = .14, RMSEA = .26. Furthermore, the two-factor model showed a significantly better fit than the single-factor model, $\Delta\chi^2(1) = 121.25, p < .001$.

Power Analysis

We conducted power analyses using G*Power (Version 3.1.9.6, Faul et al., 2007) to determine the required sample size to test our hypotheses with a power of .80. The assumed effect sizes were based on previous gamification research (e.g., Harman & Brown, 2022). The analyses for the for 2×2 mixed ANOVAs related to Hypotheses 1 and 2 revealed a necessary sample size of $N = 200$ for a small

interaction effect ($f = .10$) and a sample size of $N = 98$ for a medium-sized between-subjects effect ($f = .25$), assuming a correlation between repeated measures (i.e., pre and posttest) of $r = .50$ (see below for more information concerning the actual study design). Additionally, for the power analyses for the multivariate analyses of variance (MANOVA) related to Research Question 1, we assumed an effect size of $f^2 = .15$. This analysis revealed a sample size of $N = 92$ for five dependent variables.

Preliminary Analyses

Before running the statistical analyses, we tested whether assumptions were met. Box plots showed that there were no outliers in the data (i.e., no data point more outside than three times the interquartile range). All variables met the requirement of variance homogeneity (Levene-test: $p > .05$). Box's test was used to test the equality of covariances and following the recommendations from Verma (2015) and Warner (2012), no crucial deviations from the equality of covariances was found.

In a preliminary analysis, we tested whether the experimental groups differed regarding sex, age, highest educational degree, and video game experience. A χ^2 -test showed no significant difference for sex, $\chi^2(2) = 0.47, p = .49$, and separate ANOVAs found no significant differences for age, educational degree, and video game experience between the two groups, all F s < 1.08 , all p s $> .30$. Furthermore, as participants could take part in the study remotely or in person at the university, we also tested whether these two groups differed in their demographic variables. Again, no significant differences were found between the groups, all F s < 3.43 , all p s $> .06$; $\chi^2(1) = 1.02, p = .31$.

Table E1. Items used to measure the applicant reaction variables

Scale	Items used in the current study	Source
Clarity of work activity	I have a clear idea of what it is like to work at this company. I know which work tasks would be expected of me in this job. I have a clear idea of what the daily work routine at this company would be like.	Self-developed
Institutional image	I have always had a good impression about [name of the organization]. In my opinion, this [name of the organization] has a good image in the minds of consumers. I believe that this [name of the organization] has a better image than its competitors.	Nguyen and Leblanc (2001)
General procedural fairness	I think that this test is a fair way to select people for the apprenticeship integrated study program. I think that the test itself is fair. Overall, the method used was fair.	Bauer et al. (2001)
Job-relatedness	Doing well on this test means a person can do well in the apprenticeship integrated study program. A person who scored well on this test will also do well in the apprenticeship-integrated degree program.	Bauer et al. (2001)
Opportunity to perform	I could really show my skills and abilities through this test. This test allowed me to show what my job skills are. This test gives applicants the opportunity to show what they can really do. I was able to show what I can do on this test.	Bauer et al. (2001)
Organizational attractiveness	For me, this company would be a good place to work. I would not be interested in this company except as a last resort. This company is attractive to me as a place for employment. I am interested in learning more about this company. A job at this company is very appealing to me.	Highhouse et al. (2003)
Enjoyment	I enjoyed this test. I find this test very interesting. I find this test entertaining.	Wilde et al. (2009)

Note. Items were presented using a 5-point rating scale (1 = *strongly disagree* to 5 = *strongly agree*). All items were presented in German.

References

- Bauer, T. N., Truxillo, D. M., Sanchez, R. J., Craig, J. M., Ferrara, P., & Campion, M. A. (2001). Applicant reactions to selection: Development of the Selection Procedural Justice Scale (SPJS). *Personnel Psychology*, *54*(2), 387–419. <https://doi.org/10.1111/j.1744-6570.2001.tb00097.x>
- 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*(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Harman, J. L., & Brown, K. D. (2022). Illustrating a narrative: A test of game elements in game-like personality assessment. *International Journal of Selection and Assessment*, *30*(1), 1–10. <https://doi.org/10.1111/ijsa.12374>
- Highhouse, S., Lievens, F., & Sinar, E. F. (2003). Measuring attraction to organizations. *Educational and Psychological Measurement*, *63*(6), 986–1001. <https://doi.org/10.1177/0013164403258403>
- IBM Corp. (2021). *IBM SPSS Statistics for Mac* (Version 28.0) [Computer software]. IBM Corp.
- Nguyen, N., & Leblanc, G. (2001). Image and reputation of higher education institutions in students' retention decisions. *International Journal of Educational Management*, *15*(6), 303–311. <https://doi.org/10.1108/EUM0000000005909>
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Verma, J. P. (2015). *Repeated measures design for empirical researchers*. Wiley.
- Warner, R. M. (2012). *Applied statistics: From bivariate through multivariate techniques* (2nd ed.). Sage.
- Wilde, M., Bätz, K., Kovaleva, A., & Urhahne, D. (2009). Überprüfung einer Kurzsкала intrinsischer Motivation (KIM) [Validation of a short scale of intrinsic motivation]. *Zeitschrift für Didaktik der Naturwissenschaften*, *15*, 31–45. <https://pub.uni-bielefeld.de/record/2404161>