



rijksuniversiteit
groningen

An Examination of Applicants' Reactions and
Attitudes towards a Gamified Cognitive Ability
Test

Lea-Marie Schick

Master Thesis - Work, Organizational & Personnel Psychology

S5295076

June 2024

Department of Psychology
University of Groningen
Examiner/Daily supervisor:
Dr. Samantha Adams

A thesis is an aptitude test for students. The approval of the thesis is proof that the student has sufficient research and reporting skills to graduate but does not guarantee the quality of the research and the results of the research as such, and the thesis is therefore not necessarily suitable to be used as an academic source to refer to. If you would like to know more about the research discussed in this thesis and any publications based on it, to which you could refer, please contact the supervisor mentioned.

Abstract

This experimental study investigated applicants' perceptions of a gamified cognitive ability test compared to its traditional counterpart. This encompassed measuring process satisfaction, predictive validity, test fairness perceptions, and an examination of the impact on organizational attractiveness. Additionally, test performance's influence on the effects of gamification on applicant perceptions was explored. Contrary to initial expectations, when engaging with the traditional assessment as opposed to its gamified counterpart, higher levels of process satisfaction and perceived predictive validity, and then test fairness and organizational attractiveness were reported by applicants. Test performance served as a mediator, influencing the relationship between gamification and both process satisfaction and perceived test fairness. These findings emphasize the need for a cautious approach when implementing gamified assessments in real-world applications. A thorough evaluation, considering initial reactions and the mediating role of test performance, is crucial for informed decisions about integrating innovative assessment methods in recruitment practices.

Keywords: gamification, game-related assessment, applicant reactions, test performance, cognitive ability tests.

Table of Contents

Abstract	3
Introduction	6
Game-Related Assessments.....	8
Applicant Reactions and Attitudes towards GRAs	10
Theoretical framework	11
Gamification’s Influence on Applicant Reactions.....	12
Gamification Influences Organizational Attractiveness.....	14
Applicant Reactions and Test Performance.....	16
Method	17
Participants	17
Design and Procedure.....	19
Measures and Materials.....	19
Intervention	19
Questionnaire	22
Data analysis	23
Results	24
Hypothesis testing	24
Hypothesis 1	24
Hypothesis 2	25
Hypothesis 3	27
Hypothesis 4	27
Hypothesis 5	27

Exploratory analyses 28

Discussion 30

 Strengths and Practical Implications 34

 Limitations and Implications for Future Research 35

Conclusion..... 36

References 38

Appendix A: Fictional Application Scenario 49

Appendix B: Questionnaire 52

Appendix C: Preliminary Analyses 54

Appendix D: Detailed Results from Exploratory Analyses..... 56

An Examination of Applicants' Reactions and Attitudes towards a Gamified Cognitive Ability Test

The ability to attract, motivate and retain top talent has become a serious competitive advantage for organisations. Due to demographic change, the worlds' labor force is aging and the number of children is decreasing, i.e., fewer young people enter the workforce (Spardel, 2023). As a result, the world faces a global skilled worker shortage (OECD, 2016). In Germany alone, 400.000 new workers from non-EU countries would be needed to cover this (Banse & Buermeyer, 2023). Consequently, organizations have an increasingly hard time to fill open positions and must compete for candidates (Küpper et al., 2021). Additionally, digitalization has had a significant impact on employee hiring and now affects all phases of the recruitment and selection cycle (Woods et al., 2020). As the workforce becomes progressively digitally savvy, young talent expects digitalized processes at the organizations they apply to (PricewaterhouseCoopers, 2012). Küpper et al. (2021) propose that the use of (digital) recruiting tools that match candidates' expectations is one way of winning over talent.

Online testing has traditionally been the focus of research into how technology affects employee selection (Ryan & Ployhart, 2014). As reported by Nikolaou (2021), two new selection techniques have now drawn more interest: asynchronous interviews and game-related assessments. Asynchronous interviews are a sort of interview where candidates are asked to record their responses to a series of interview questions and submit them online (Nikolaou, 2021). On the other hand, Game-Related Assessments (GRA) are based on games (Ramos-Villagrasa et al., 2022). Some GRAs are based on gamification, the inclusion of game elements to non-gaming contexts (Georgiou et al., 2019). Selection methods can be gamified by adapting or digitalizing existing selection methods, like situational judgment tests or cognitive ability tests (Nikolaou, 2021). For example, Hommel et al. (2022) created the Gamified Set-Shifting Task (GSST) based on the Wisconsin Card Sorting Test (WCST). They

included a work-specific simulation where participants take the role of a marketing director and added game design elements such as storytelling, collecting points and a performance graph. Georgiou et al. (2019) gamefully designed a situational judgment test measuring for example resilience, decision-making, flexibility, and adaptability. This GRA involves selecting an avatar with a backstory, storytelling, rewards, and feedback. Previous research has shown that GRAs can be reliable and valid, as well as leading to positive applicant reactions and increased organizational attractiveness (Georgiou & Nikolaou, 2020). Furthermore, they are seemingly appealing to younger applicants (Nikolaou, 2021). Studies like Georgiou (2021) and Georgiou and Lievens (2022), however, indicate that research should keep looking into the fundamental mechanisms and strategies for enhancing the GRA-reactions link, as well as broaden the research to various assessments and more applicant reaction variables (Georgiou & Nikolaou, 2020).

Among the most researched facets of applicant reactions are fairness perceptions and predictive validity perceptions (Chan & Schmitt, 2004). Ryan and Ployhart (2000) describe applicant reactions as the “attitudes, affect or cognitions an individual might have about the hiring process” (p. 566). According to signalling theory, candidates can use information gathered through their experiences during selection assessments as signals for the organization and job they are applying for (Spence, 1973). If these perceptions are positive and pleasant, this influences the organization’s attractiveness that may lead to further positive outcomes for the organization.

Landers and Sanchez (2022) postulate that, as with game design, the success of gamification in personnel selection depends on how well game mechanisms are incorporated into the current system. Adding game mechanics does not automatically make an evaluation more interesting or fun. Taken together with the reasons mentioned above and given the early stage of gamification research, examining applicants’ reactions should be of high relevance for organizations and needs to be further researched. Specifically, additional research is

required to compare gamified and conventional selection processes (Aouam et al., 2023). If applicants see a selection method used by a certain organization as more favourable, they are more likely to accept a job offer. Furthermore, they are then more likely to recommend the firm to their network, increasing the number of future applications (Ployhart & Harold, 2004). In contrast, negative candidate experience can lead to never applying at the respective firm again (Baker, 2018). This is especially important to consider in the context of skilled-worker shortage.

The purpose of this study is to broaden previous research about applicant reactions and attitudes towards gamification onto a gamified cognitive ability assessment. Specifically, applicant reactions to the WCST and Hommel et al.'s (2022) GSST, which both assess cognitive flexibility, are thoroughly compared. I therefore aimed to reduce the current literature gap about whether previous findings, e.g. from Georgiou and Nikolaou (2020), can be generalized on other GRAs, different application scenarios and samples. Moreover, this study explored the currently unknown influence of test performance on the gamification-applicant reactions link, thus narrowing the gap about the fundamental principles of this relationship.

In the following sections, I will provide an overview of the current research on game-related assessments, including psychometric properties and applicant reactions to support the development of the hypotheses for this study.

Game-Related Assessments

Based on Landers and Sanchez' (2022) categorization, Ramos-Villagrasa et al. (2022) proposed a more detailed, continuous classification of GRAs with traditional assessments on one end and playful games created for fun at the other end of the spectrum. They distinguish at least four types of GRAs: (1) Gamified assessments are based on already existing assessments and were redesigned using gamification, i.e., by adding individual game elements. One example is Hommel et al.'s (2022) *Gamified Set-Shifting Task*, a modification

of the Wisconsin Card Sorting Test, which were assessed in this study. (2) Gamefully designed assessments are completely new assessments that were designed using individual game elements. An example is *Owiwi*, a situational judgement test by Georgiou et al. (2019). (3) Game-based assessments (GBA), also serious games, measure job-related constructs through games without entertainment purpose, e.g., *Virus Slayer* (Wiernik et al., 2022) tests candidates for cyber occupations in the US Air Force. Dozens or hundreds of game mechanics are integrated and the player experiences them in a core gameplay loop (Landers & Sanchez, 2022). Lastly, (4) Playful games used for assessment purposes are conventional games used to gather information about specific abilities, e.g., Sanchez et al. (2022) used the two commercial off-the-shelf (COTS) games *Project M* and *Richie's Plank Experience* to assess emotional intelligence. It is proposed that the closer an assessment is to the playfulness extreme, the more positive applicants' reactions will be, such as organizational attractiveness (Gkorezis et al., 2011, Ramos-Villagrasa et al., 2022).

However, before recommending GRAs, research should provide empirical evidence to support their use. GRAs must meet the following psychometric standards: (1) acceptable reliability so that measurement is accurate, (2) construct validity to ensure GRAs measure what is meant to be measured (3) predictive validity to predict the criterion of interest (usually job performance), and (4) freedom from bias to ensure scores are not influenced by personal characteristics, such as gender or age (Landers et al., 2022; Salgado et al., 2017; Wiernik et al., 2022). In addition, GRAs should positively influence applicant reactions (Ramos-Villagrasa et al., 2022).

While the exact prevalence of GRAs in organizations is unknown, large companies like Google, PwC and L'Oréal have started to use gamified tools for recruitment (Buil et al., 2020). Moreover, GRAs have been of interest for researchers in the past years. In a systematic literature review about GRAs, Ramos-Villagrasa et al. (2022) identified 34 scientific articles published between 2012 and 2022, of which 14 were published in 2022 alone.

Previous studies have mixed results concerning GRAs' psychometrics. Landers et al. (2022) observed convergence between performance in a cognitive ability GBA and in a cognitive ability test battery. Nonetheless, they also noted gender differences in the game scores, endangering the freedom from bias, essential in personnel selection. On the other hand, Wu et al. (2022) found that their two GBAs, which were meant to assess conscientiousness, instead evaluate cognitive ability and the remaining factors of the Big Five. Regarding predictive validity, results are more positive. Various studies have found associations between GRAs and academic performance (e.g., Auer et al., 2022; Hommel et al., 2022), and supervisory ratings (Landers et al., 2022). However, Landers and Sanchez (2022) state that the validities of gamified assessments need their own evidence. Phrased differently, findings from a particular GRA cannot unconditionally be generalized to another one. Nevertheless, based on their systematic review, Ramos-Villagrasa et al. (2022) conclude that GRAs can generally be utilized in personnel selection, although the benefits over traditional assessments are less substantial than anticipated: At this stage of research, GRAs only offer enough advantages over conventional techniques if the aim is to enhance applicant reactions during the selection process.

Applicant Reactions and Attitudes towards GRAs

Applicant reactions have been referred to as "attitudes, affect, or cognitions an individual might have about the hiring process" (Ryan & Ployhart, 2000, p. 566). According to Hausknecht et al. (2004), there are various reasons for researching applicants' reactions. Firstly, maintaining a positive company image during the selection process is critical since losing top candidates costs money (Murphy, 1986). Applicants who find some components of the selection system intrusive may regard the organization as a less appealing alternative during the job search process. Second, candidates who have unfavorable reactions to a selection process may discourage other potential applicants from applying for jobs with the company (Smither et al., 1993). Similarly, candidates with positive reactions are more likely

to recommend the organization to others (Hausknecht et al., 2004). Third, candidates may be less inclined to accept an offer from a company with unfavorable selection methods (Macan et al., 1994). For instance, Gilliland (1993) argued that applicants who see injustice during the selection process may decline job offers out of concern that unfair recruitment practices indicate that the company treats its employees badly. Fourth, applicant reactions may be linked to legal complaints and court challenges, as applicants who believe a specific selection technique is invasive or improper may be more likely to file a lawsuit than applicants who believe the process is fair and face valid (Smither et al., 1993). Lastly, applicants who feel mistreated throughout the selection process are less inclined to reapply with an organization or purchase the company's products or intend to do so (e.g., Ployhart & Ryan, 1998). Bad candidate experience cost Virgin Media at least \$5M annually due to cancelled subscriptions of candidates and their network (Steiner, 2017).

Theoretical framework

Gilliland's (1993) organizational justice framework has been the most important theoretical framework in applicant reactions (Truxillo et al., 2016). It involves the perceived fairness of outcome allocations (distributive justice), the rules and processes used to arrive at those decisions (procedural justice), the consideration and respect offered to individuals (interpersonal justice), and the justifications and accounts provided to them (informational justice) (Greenberg, 1993; Hausknecht et al., 2004). Applicants' perceptions of these facets then influence future attitudes, self-perceptions, and behaviours, including outcomes like hiring choices (Smither et al., 1993). As an example, candidates who feel they received unfair treatment during an interview may be less likely to accept a job offer or recommend the company to others. Gilliland (1993) emphasised the role of procedural organizational justice, which has applications to technology as well (Nikolaou et al., 2019). For instance, technology, particularly online testing, has made it possible for hiring organizations to provide quick feedback to candidates, while distancing the applicant and the company, which has

consequences for how applicants experience the pillar interpersonal treatment of Gilliland's (1993) framework (Sears et al., 2013).

Hausknecht et al. (2004) extended Gilliland's (1993) framework and an adaption of Ryan and Ployhart (2000) work, and proposed a more recent model of applicant reactions to selection processes. The model's fundamental concept is that applicant views of the selection process can best predict significant outcomes, such as organizational attractiveness (Macan et al., 1994) and reapplication intentions (Ployhardt & Ryan, 1998). Applicant perceptions include their opinions on the various aspects of organizational justice, their sentiments towards testing, and their overall attitudes toward tests and selection. In addition, four large categories of antecedent variables are suggested as determinants of applicant views. These consist of person characteristics, perceived procedure characteristics, job characteristics, and factors associated with the organizational context. Perceived procedure characteristics and applicant perceptions differ in their conceptual meaning such that perceived procedure characteristics refer to elements specific to the selection process or procedures, while applicant perceptions are more general assessments of the process (Hausknecht et al., 2004).

Based on these theoretical frameworks, the expected relationships and hypotheses will be argued in the subsequent sections.

Gamification's Influence on Applicant Reactions

In their systematic literature review, Ramos-Villagrasa et al. (2022) conclude that GRAs tend to be more highly regarded than traditional selection methods (e.g., Collmus & Landers, 2019; Harman & Brown, 2022; Landers et al., 2021) and encourage positive applicant reactions (al-Qallawi & Raghavan, 2022; Georgiou, 2021). These results appear to be consistent among various types of GRA, given that these studies have been undertaken with a wide variety of games, ranging from standard assessments to include serious games.

Among the most researched facets of applicant reactions are fairness perceptions, which are explained above, and predictive validity perceptions (Chan & Schmitt, 2004).

Perceived predictive validity is defined as perceptions about “how well the procedure predicts future job performance, regardless of how it looks” (Smither et al., 1993, p. 54). Instead of being a psychometric quality, perceived predictive validity is a person’s assessment of how well a selection technique predicts outcomes. To date, there is not much research on how gamification of assessments influences applicants’ test fairness perceptions. However, applicant reactions literature (e.g., Bauer et al., 2001; Gilliland, 1993; Hausknecht, et al., 2004) shows that perceived predictive validity predicts a method's perceived fairness. Georgiou and Nikolaou (2020) examined whether candidates' perceptions of test fairness towards a gamefully designed Situational Judgement Test are mediated by predictive validity. They found no variations between the fairness perceptions of the gamefully designed and traditional assessment method and no mediation through perceived predictive validity. On the other hand, Hommel et al. (2022) compared the acceptance of a gamified cognitive ability test with its traditional version. Perceived predictive validity and procedural justice perceptions were part of this acceptance scale, and the GRA had higher acceptance scores than the traditional version. Since the aforementioned studies had mixed results, more research is needed to clarify gamification’s influence.

In addition to perceived validity, Smither et al. (1993) inspected emotion (i.e., how much candidates enjoyed the test), connected to procedural justice, or the fairness of the selection processes. The theoretical model of organizational justice by Barsky et al. (2011, p. 271) states that "affect creates the context through which people experience, appraise, and draw conclusions in matters of fairness." This means the inferences applicants make about the fairness of the selection process may be influenced by how satisfied they feel with the selection process. Since gamification has been defined as a new and more enjoyable approach to recruiting, the degree to which candidates are satisfied with the selection process may be higher for gamified assessments compared to traditional ones (Collmus et al., 2016). In fact, it has been shown that calling an ability test a "game" enhances applicant responses by giving

the impression that time is passing quickly (Collmus & Landers, 2019). In Georgiou and Nikolaou's (2020) study, applicants' process satisfaction fully mediated the effect of assessment method on perceived fairness. However, the impact of gamification on perceived test fairness through enjoyment or satisfaction that candidates feel from the selection process has not yet been tested with regards to a cognitive ability test, namely Hommel et al.'s (2022) GSST.

Based on Hausknecht et al.'s (2004) theoretical framework, I hypothesize that the cognitive ability test enriched with a game fiction component may have a favorable impact on candidates' satisfaction with the selection process and perceptions of its predictive validity, and that these positive perceptions may generate more positive judgments about the test's fairness.

***Hypothesis 1:** Participants completing the GRA will have higher levels of a) perceived predictive validity, b) process satisfaction and c) perceived test fairness.*

***Hypothesis 2:** The influence of assessment type on perceived test fairness will be mediated by two factors: a) perceived predictive validity and b) process satisfaction.*

Gamification Influences Organizational Attractiveness

For businesses to succeed, they must be seen as desirable places to work. Recruitment and selection procedures are crucial if a company wants to attract top talent. To positively affect candidates' opinions of the selection process and the recruiting organization, it is crucial to keep candidates engaged and motivated (Georgiou & Lievens, 2022). Game-like experiences are enjoyable, motivating, and engaging. Signaling theory postulates that candidates who get unclear or incomplete information utilize this information as "signals" about the position they are applying for and the organizational characteristics (Spence, 1973). Symbolic traits are the less obvious characteristics of a company (its "personality") that job applicants may gather. People are drawn to these traits, for example competence or

innovation, to convey their ideals or to impress others (Highhouse et al., 2007). They contribute significantly candidates' perceptions of an employer, including their attraction towards the company (Lievens & Highhouse, 2003). Moreover, they forecast organizational attractiveness in connection to hiring results beyond instrumental qualities (Lievens & Highhouse, 2003). Like other selection processes, the usage of a gamified assessment approach throughout the selection process may "signal" to candidates about the organization's values and characteristics. After completing a gamified assessment method, candidates may form positive perceptions of predictive validity, process satisfaction, and ultimately test fairness. These perceptions may indicate to them that the organization they are applying to is a fair organization, increasing the company's attractiveness and the likelihood of a successful outcome. It has been demonstrated that applicant reactions, such as organizational attractiveness, and the use of new technology in selection are mediated by views of procedural justice (e.g., Bauer et al., 2006). Georgiou and Nikolaou (2020) found a significant effect of assessment method on organizational attractiveness through job satisfaction and perceived fairness, but not perceived predictive validity. According to Georgiou and Lievens (2022), the relationship between assessment method and organizational attractiveness was mediated by the enjoyment and flow of GRAs that led candidates to regard the company as more innovative and competent. However, more studies with different groups of people are required. Based on the theoretical framework presented above, I suggest that a GRA, in comparison to the traditional version of the assessment, will increase applicants' satisfaction with the process and perceptions of the test's fairness, which will then increase the organization's perceived attractiveness to candidates.

***Hypothesis 3:** Participants completing the GRA will demonstrate greater organizational attractiveness levels than participants completing the traditional assessment.*

Hypothesis 4: Assessment type first influences organizational attractiveness through a) participants' perceptions of predictive validity and b) process satisfaction, and then through perceived test fairness.

Applicant Reactions and Test Performance

It is possible that applicants' test performance affects perceptions they have after taking a selection test because of self-serving bias (Chan et al., 1998; Ployhart & Harold, 2004). Self-serving bias is an individual's tendency to attribute success to one's own personal characteristics, while failure is attributed to external factors (Miller, 1978). Fiske and Taylor (1991) claim that unexpected, stressful, unusual, negative, and significant occurrences cause someone to employ attributional heuristics like the self-serving bias. Since such situations are common in selection contexts, attributional heuristics are relevant here (Ployhart & Ryan, 1997). Multiple researchers have found small to moderate positive associations between applicant perceptions and actual and perceived selection test performance (e.g., Chan & Schmitt, 1997; Macan et al., 1994; Oostrom et al., 2012; Schmit & Ryan, 1997; Smither et al., 1993). For instance, Hausknecht et al. (2004) found small correlations between actual procedure performance and perceived predictive validity, procedural justice perceptions, and organizational attractiveness. The correlations with perceived procedure performance were even moderate to large.

Presently, research examining the influence of test performance on applicant reactions to GRAs is scant. In both test versions used in this study, applicants received feedback on their performance. However, because of the gamified nature of the test, more feedback was given for applicants in the gamified version. A performance graph and game score were displayed on the screen during the game. Participants also received a visual notification of an increase vs. decrease of their score for each correct vs. incorrect match, and a fanfare sound was played for correct matches. At the end of the test, the total game score and profit was

displayed. Applicants in the control condition only received a correct vs. false feedback after matching. By adding feedback through scoring, a common element in game design (Landers & Sanchez, 2022), participants in the gamified condition compared to the non-gamified condition should be more aware of their test performance. Phrased differently, test performance should be more salient in the gamified condition.

Due to the difference in the salience of performance, applicant reactions may be more pronounced in the gamified condition than those in the traditional condition. Consequently, I hypothesize that:

***Hypothesis 5:** Participants' test performance will moderate the effect of assessment method on the aforementioned applicant reactions to the selection procedure, and moderation will be stronger for the gamified test version.*

Method

Participants

Participants were recruited through Prolific, a research-focused panel company. Prior research has demonstrated that data collected from Prolific is of high quality, compared to data obtained from other panel companies and university subject pools (Palan & Schitter, 2018; Peer et al., 2017). Additionally, participants were recruited via my personal network and approached through social media networks (WhatsApp, Instagram, LinkedIn). All participants had to be German speakers, as the Gamified Cognitive Ability Test is only available in German. Respondents from Prolific were compensated with 6€/hour. Those recruited through personal contacts were not compensated.

Data was collected from July to September 2023. A total of 565 people started the Qualtrics Survey. 15 participants did not consent to the processing of their data at the beginning of the survey, 157 did not complete the survey and 2 withdrew their consent to data usage after the debriefing, resulting in a dropout rate of 30.8%. Chi-square tests of

independence were conducted to examine whether dropout was related to the assigned condition. When comparing the combined sample and the sample from the researcher's network, the association was not statistically significant (combined: $\chi^2(df = 1) = 2.74, p = .098$; personal contacts: $\chi^2(df = 1) = 0.56, p = .45$). For the Prolific sample, the association was statistically significant ($\chi^2(df = 1) = 4.48, p = .034$), suggesting that participants in the gamified condition were more likely to drop out of the study. The final sample consisted of $N = 391$ participants (198 female, 186 male) (see Table 1). Their age ranged from 18 to 79 years

Table 1*Sociodemographic Characteristics of Participants*

Characteristic	Prolific		Personal contacts		Full sample	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Age						
18-34	180	72.0	105	74.5	285	72.9
35-49	54	21.6	11	7.8	65	16.6
≥ 50	16	6.4	24	17.0	40	10.2
Did not specify	0	0	1	0.7	1	0.3
Gender						
Female	118	47.2	80	56.7	198	50.6
Male	125	50.0	61	43.3	186	47.6
Non-binary	5	2.0	0	0	5	1.3
Did not specify	2	0.8	0	0	2	.5
Highest educational level						
Lower than High school	21	8.4	6	4.2	27	7.0
High school ("Abitur")	49	19.6	19	13.5	68	17.8
Vocational training	13	5.2	19	13.5	32	8.2
Bachelor or equivalent	85	34.0	51	36.2	136	34.8
Master or equivalent	76	30.4	44	31.2	120	30.7
PhD	6	2.4	2	1.4	8	2.0
Country of Residence						
Germany	103	41.2	117	83.0	220	56.3
Poland	41	16.4	0	0	41	10.5
Netherlands	10	4.0	16	11.3	26	6.6
UK	19	7.6	1	0.7	20	5.1
Portugal	12	4.8	0	0	12	3.1
Greece	11	4.4	0	0	11	2.8
Austria	10	4.0	1	0.7	11	2.8
Other	44	17.6	6	4.3	50	12.8

Note. $N = 391$ ($n_{Prolific} = 250$; $n_{Network} = 141$).

($M = 31.53$, $SD = 11.18$, $N = 390$). 67.5% of participants had a bachelor degree or higher educational background. Participants lived in various countries, but most currently lived in Germany (56.3%).

Design and Procedure

To test the respective hypotheses, an experimental study was conducted. Participants received a link to a Qualtrics survey. After being informed about the study and asked for consent to participate, they were presented with a fictional application scenario (see Appendix A. Participants were asked to imagine they are applying for a Marketing Director position at a hypothetical company and have been invited to participate in a selection assessment. Next, participants were randomly allocated to one of two conditions. The experimental group completed the GSST, a gamified cognitive ability test (Hommel et al., 2022). The control group completed an equivalent computer-version of the WCST (Heaton et al., 1993). Once they had completed the assessment, participants answered a questionnaire including measures for perceived test fairness, organizational attractiveness, predictive validity, process satisfaction and questions about their demographics. The survey could only be completed on a laptop, as the GSST does not support touch gestures and therefore does not work on mobile devices and tablets. At the end, respondents were debriefed by receiving detailed information about the study, including the experimental conditions.

Measures and Materials

Intervention

Two experimental conditions were created using the same assessments used in Hommel et al.'s (2022) study.

Wisconsin Card Sorting Test. The WCST assesses cognitive flexibility, the ability to adjust cognitive processing strategies according to changing, unexpected and new circumstances (Cañas et al., 2003). Participants were prompted to match a series of response cards with one of four reference cards. These were shown at the top of the screen against a

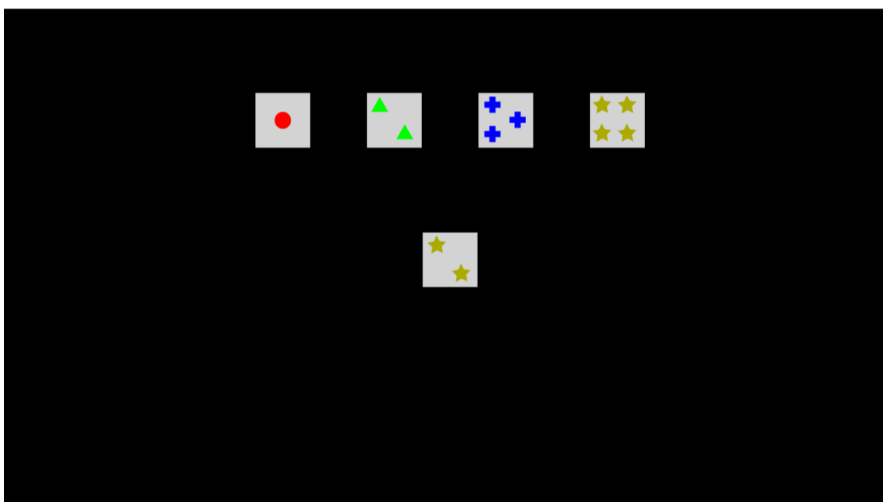
black background (see Figure 1). The response cards were displayed one at a time at the bottom of the screen. The proper sorting criterion, i.e., object count, colour, or shape, had to be inferred by the participants through trial and error based on the feedback (correct/incorrect) they got after each trial. The sorting criteria abruptly changed after 10 straight correctly chosen cards, and participants had to learn and use the new rule.

The total number of correct trials made up the test's final score. This is an indicator of one's capacity to recognize and follow a sorting rule as well as one's adaptability to changing rules as the sorting criteria do. A participant's score was reduced for unsuccessful efforts by one point, without the participant being aware of this. Higher scores therefore reflect greater degrees of cognitive flexibility.

A modification of the WCST from PsyToolkit (Stoet, 2017; www.psychtoolkit.org) was used in this study to make it comparable to the original test (Heaton et al., 1993). The number of trials was increased to 64, and the selection criteria were adapted so that the sorting rule changed after the test-taker had completed 10 consecutive correct trials. In comparison, the PsyToolkit version of the WCST only has 60 trials and changes the classification rule after 10 cards.

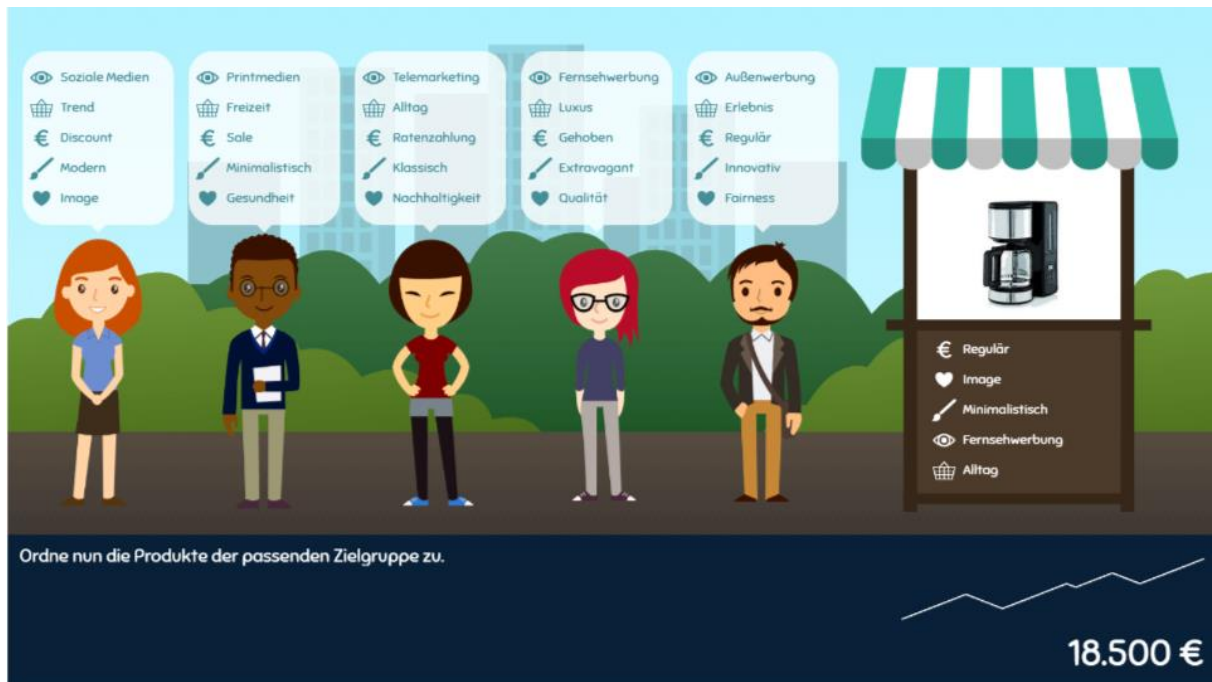
Figure 1

Interface of the Wisconsin Card Sorting Test



Gamified Set-Shifting Task. The GSST adheres to the same basic concepts as the original WCST, but contains a higher level of difficulty, a work-related design and game components such as points, a performance graph, and storytelling. Instead of matching cards based on shape, colour, and object count, participants were required to match products to target groups based on common characteristics.

The game began with a fictional marketing agency employee congratulating the player on their new employment as a Marketing Director. The fictional employee informed the test-taker about the goal of implementing a new marketing strategy to lower expenses and boost the effectiveness of marketing campaigns. Then, the game interface was introduced (see Figure 2). On the left side of the screen, five target groups were represented by five avatars, and consumer products were displayed one at a time on a market stand at the left side of the screen. Test-takers started with a budget of €10,000. They were instructed to match products to the target groups based on the common characteristics (communication channel, product category, pricing, product design, values). The participant was not informed of the proper matching rule. Until the target characteristic changed, each target group was open to one aspect of the consumer product. Participants had to deduce which of the characteristics the target groups were now responsive to based on the feedback they received after each trial: A visual and audio reward cue (an erupting confetti cannon and the sound of a fanfare) was played each time a product was matched in accordance with the current rule. Furthermore, the participant was informed their balance had increased by €500. For unsuccessful trials, the players' score was reduced by €500. The fundamental sorting rule changed after seven successful matches in a row, and the player had to adjust by giving up the old strategy and adopting a new one. A performance graph displaying the progress over earlier trials was shown at the bottom right of the screen along with the current game score and account balance.

Figure 2*Interface of the Gamified Set-Shifting Task*

The final game score was shown after all 64 trials had been completed. Like the WCST, it was calculated based on the sum of all successful and unsuccessful attempts, operationalizing the capacity to recognize and adapt to rule changes. The GSST and WCST both use the same scoring methodology, but different scaling.

Questionnaire

To measure applicants' reactions and attitudes towards the respective assessment, we used the same scales as Georgiou and Nikolaou (2020). For this study, the original items of the selected scales were translated into German with the standard back-translation technique (Brislin, 1980) and adapted to a gender-inclusive language. Both the original, English items and the translated, German items of all scales can be found in Appendix B.

The participants were presented with the following instructions: "On this page, you will find a series of statements that relate to your experience with the technology-based assessment you just completed. Please read each statement carefully and indicate the extent to

which you agree.” All answers were rated on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree).

For the assessment of participants’ *perceived test fairness*, Kluger and Rothstein’s (1993) three items of the test fairness scale was used. A sample item is “I think the test is fair”. In the current study, Cronbach’s alpha was $\alpha = .78$.

Participants’ *predictive validity perceptions* were measured using Chan et al.’s (1998) three items of predictive validity perceptions. A sample item is “I am confident that the test can predict how well an applicant will perform on the job”. Cronbach’s alpha for this scale was $\alpha = .83$ in the current study.

Process satisfaction was assessed using Sylva and Mol’s (2009) item of overall process satisfaction: “Overall, I was satisfied with this application process.”

Organizational attractiveness was examined using the General Attractiveness Scale by Highhouse et al. (2003) including 5 items, e.g., “For me, this company would be a good place to work”. In the current study, Cronbach’s alpha for this scale was $\alpha = .93$.

Moreover, participants were asked about their demographics, including age, gender, country of residence, and highest level of education.

Data analysis

To examine Hypothesis 1, a Multivariate Analysis of Variance (MANOVA) was conducted in SPSS. Subsequently, follow-up univariate Analysis of Variances (ANOVAs) were performed for each dependent variable. Hypothesis 3 was investigated using a one-sided t-test for independent samples.

The mediation analyses were conducted using PROCESS macros for SPSS (Hayes, 2018). Specifically, Model 4 of Hayes (2018) PROCESS macro v3.4 was used to test Hypothesis 2, while Model 6 was used to explore Hypothesis 4. To examine Hypothesis 5, moderation analyses were conducted using Model 1 of Hayes (2018) PROCESS macros. Following Davidson and MacKinnon (1993), bootstrapping with 5000 samples and

heteroscedasticity consistent standard errors were utilized to calculate confidence intervals and inferential statistics. As Process macro only tests two-sided by default, the confidence interval was set to 90% for the mediation analyses to construct a one-sided 95% confidence interval. An indirect effect was considered significant if there was no zero between the lower confidence interval limit and plus infinity. For assumption testing, see Appendix C.

Results

Descriptive statistics and correlations of the study variables can be found in Table 2. The four applicant reaction variables had medium to large positive correlations with each other. Test performance was also negatively correlated with age.

Hypothesis testing

Hypothesis 1

Hypothesis 1 predicted participants completing the GRA to show higher levels in the applicant reaction variables compared to participants completing the traditional version. The

Table 2

Descriptive Statistics and Correlations of Study Variables

Variable	<i>n</i>	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7
1. Gender	391	-	-	—						
2. Age	390	31.53	11.18	.07	—					
3. Perceived predictive validity	391	2.15	0.88	.09	.04	—				
4. Process satisfaction	391	2.51	1.11	.07	.02	.60**	—			
5. Perceived test fairness	391	2.70	0.84	.06	-.06	.68**	.61**	—		
6. Organizational attractiveness	391	2.99	0.93	.05	-.05	.34**	.45**	.45**	—	
7. Test performance	391	12.70	32.60	.07	-.20**	.14**	.23**	.43**	.23**	—

Note: Two-tailed *p*-values; * *p* < .05. ** *p* < .01.

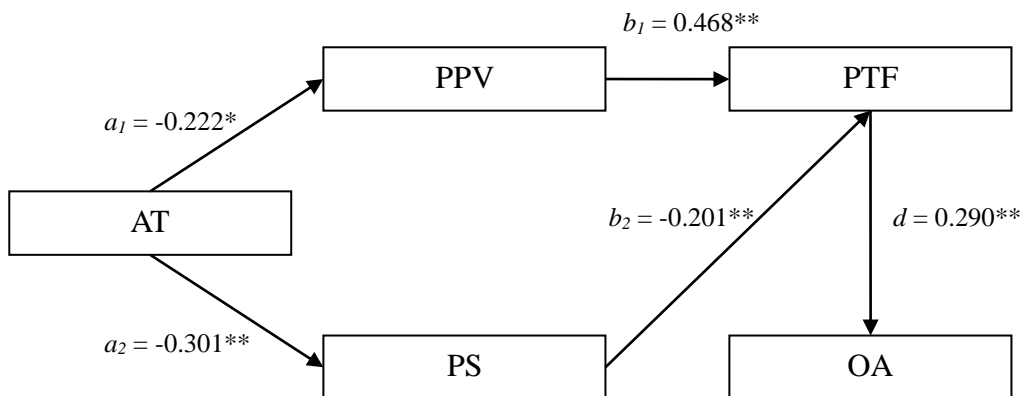
one-way MANOVA showed a statistically significant difference between the gamified and traditional assessment on the combined dependent variables, $F(3, 387) = 23.762, p < .001$, partial $\eta^2 = .156$, Wilk's $\Lambda = .844$. Therefore, post-hoc univariate ANOVAs were conducted for every dependent variable. Results showed statistically significant differences between the assessment types for perceived test fairness ($F(1,389) = 56.879, p < .001$, partial $\eta^2 = .128$), perceived predictive validity ($F(1,389) = 6.210, p < .05$, partial $\eta^2 = .016$) and process satisfaction ($F(1,389) = 517.563, p < .001$, partial $\eta^2 = .043$), in that participants in the traditional assessment scored higher than participants completing the GRA. Consequently, Hypothesis 1 was rejected.

Hypothesis 2

Hypothesis 2 predicted that the influence of the assessment type on perceived test fairness would be mediated by perceived predictive validity and process satisfaction, where participants completing the GRA would show higher levels. The mediation analysis revealed that there was no zero in the confidence intervals of the indirect effect of assessment type on perceived fairness through process satisfaction ($a_2b_2 = -0.093; SE = 0.028; 90\% CI [-0.142; -0.051]$), as well as through perceived predictive validity ($a_1b_1 = -0.104; SE = 0.041; 90\% CI [-0.174; -0.039]$), and participants completing the traditional assessment scored higher than participants completing the GRA. These indirect effects were thus considered as significant, but in the opposite direction as hypothesized. The direct effect of assessment type on test fairness perceptions was significant in the presence of the mediators ($c' = -0.402, p < .001$). Hence, both perceived predictive validity and process satisfaction partially mediated the relationship between assessment type and test fairness perceptions, but not in the hypothesized way. Detailed results of the mediation analyses can be seen in Figure 3 and Table 3. Therefore, Hypothesis 2 was rejected.

Figure 3

Mediation Model Assessment Type on Organizational Attractiveness



Note: Two-tailed p -values; * $p < .05$. ** $p < .01$. Unstandardized regression coefficients reported. Abbreviations: Assessment type (AT), Task Performance (Score), Perceived Predictive Validity (PPV), Process Satisfaction (PS), Perceived Test Fairness (PTF), Organizational Attractiveness (OA)

Table 3

Results of Regression Analyses

Y	Predictors	B	SE	t	p (two-tailed)	90% CI
PPV	AT	-0.22	0.09	-2.51	< .001**	-0.37, -0.08
PS	AT	-0.30	0.09	-3.31	< .001**	-0.45, -0.15
PTF	AT (total effect)	-0.60	0.08	-7.51	< .001**	-0.73, -0.47
	AT (direct effect)	-0.40	0.06	-6.79	< .001**	-0.50, -0.31
	PPV	0.47	0.04	10.73	< .001**	0.40, 0.54
	PS	0.20	0.04	5.68	< .001**	0.14, 0.26
OA	AT (total effect)	-0.42	0.09	-4.56	< .001**	-0.57, -0.27
	AT (direct effect)	-0.14	0.10	-1.47	.142	-0.30, 0.02
	PTV	0.29	0.08	3.45	< .001**	0.15, 0.43
	PS	0.25	0.06	4.41	< .001**	0.15, 0.34

Note: Two-tailed p -values; * $p < .05$. ** $p < .01$. Unstandardized regression coefficients reported. Bootstrap sample size 5000. Abbreviations: Assessment type (AT), Perceived Predictive Validity (PPV), Process Satisfaction (PS), Perceived Test Fairness (PTF), Organizational Attractiveness (OA)

Hypothesis 3

Hypothesis 3 predicted participants completing the GRA to have higher levels in organizational attractiveness. The t-test for independent samples showed that organizational attractiveness was lower in the gamified ($M = 2.74$, $SD = 0.92$, $n = 186$) than in the traditional condition, ($M = 3.16$, $SD = .90$, $n = 205$). This difference was statistically significant, $t(389) = 4.57$, $p < .001$, $d = 0.462$. Consequently, Hypothesis 3 was rejected.

Hypothesis 4

Hypothesis 4 predicted that participants' perceptions of predictive validity and process satisfaction, and then perceived test fairness would mediate the effect of assessment type on organizational attractiveness, where participants completing the GRA showed higher levels. The results of the mediation analysis revealed that there was no zero in the confidence intervals of the indirect effect of assessment type on organizational attractiveness through process satisfaction and then perceived test fairness ($a_2b_2d = -0.030$; $SE = 0.015$; 90% CI [-0.57; -0.009]), as well as through perceived predictive validity and then perceived test fairness ($a_1b_1d = -0.018$; $SE = 0.008$; 90% CI [-0.033; -0.006]), and participants completing the traditional assessment scored higher than participants completing the GRA. These indirect effects were thus considered as significant, but in the opposite direction as hypothesized. Furthermore, the direct effect of assessment type on organizational attractiveness was not significant in presence of the mediators ($c' = -0.139$, $p = .142$). Hence, both perceived predictive validity and process satisfaction and then perceived test fairness fully mediated the relationship between assessment type and test fairness perceptions, but not as hypothesized. Therefore, Hypothesis 4 was rejected.

Hypothesis 5

Moderation analyses were run to determine whether the interaction between assessment type and task performance predicts perceived predictive validity, process satisfaction, perceived test fairness, and organizational attractiveness. Detailed results of the

analyses can be seen in Table 4. As task performance did not significantly moderate the effects between assessment type and any of the reaction variables, hypothesis 5 was rejected. However, with task performance added to the model, the independent effects of assessment type on the reaction variables all were non-significant. In addition, the independent effect of task performance on perceived test fairness was significant ($p = .004$).

Exploratory analyses

As the results from testing for hypothesis 5 suggest that task performance is somehow involved in the effect of assessment type on the applicant reaction variables, but not as a

Table 4

Results of Moderation Analyses with Task Performance as Moderator

Y	Predictors	B	SE	t	p (two-tailed)	95% CI
PPV	AT	0.02	0.17	0.11	.915	-0.31, 0.35
	Score	0.01	0.01	1.43	.153	-0.01, 0.03
	AT * Score	-0.01	0.01	-1.08	.280	-0.01, 0.01
Overall model summary: $F(3, 387) = 3.02, p = .030, R^2 = .024$						
PS	AT	-0.20	0.22	-0.90	.367	-0.63, 0.23
	Score	0.01	0.01	1.05	.294	-0.01, 0.03
	AT * Score	< -0.01	0.01	-0.55	.585	-0.02, 0.01
Overall model summary: $F(3, 387) = 7.52, p < .001, R^2 = .054$						
PTF	AT	-0.11	0.15	-0.76	.450	-0.41, 0.18
	Score	0.02	0.01	2.94	.004*	0.01, 0.04
	AT * Score	-0.01	< 0.01	-1.57	.118	-0.02, 0.01
Overall model summary: $F(3, 387) = 29.83, p < .001, R^2 = .189$						
OA	AT	-0.29	0.16	-1.81	.071	-0.61, 0.03
	Score	0.01	0.01	0.58	.563	-0.01, 0.02
	AT * Score	< -0.01	0.01	-0.08	.938	-0.01, 0.01
Overall model summary: $F(3, 387) = 7.79, p > .001, R^2 = .058$						

Note: Two-tailed p -values; * $p < .05$. ** $p < .01$. Unstandardized regression coefficients

reported. Bootstrap sample size 5000. Abbreviations: Assessment type (AT), Task

Performance (Score), Perceived Predictive Validity (PPV), Process Satisfaction (PS),

Perceived Test Fairness (PTF), Organizational Attractiveness (OA)

moderator, I tested whether task performance served as mediator between assessment type and the applicant reaction variables. PROCESS macros v3.4 Model 4 for SPSS was used to perform simple mediation analyses (Hayes, 2018) like testing for hypothesis 2. Detailed results can be seen in Table 5.

Results showed that the relationship between assessment type and process satisfaction was fully mediated by test performance, indirect effect $ab = -0.286$, $SE = 0.138$, 95%-CI [-0.563, -0.016]. The effect of assessment type on perceived test fairness was also shown to be fully mediated by test performance, indirect effect $ab = -0.503$, $SE = 0.101$, 95%-CI [-0.699, -0.302]. Test performance did not mediate the relationships between assessment type on

Table 5

Results of Mediation Analyses with Task Performance as Mediator

Y	Predictors	B	SE	t	p (two-tailed)	95% CI
PPV	AT (direct effect)	-0.56	0.14	-0.41	.680	-0.28, 0.17
	AT (total effect)	-0.22	0.09	-2.51	.013*	-0.40, -0.05
	Score	< 0.01	< 0.01	1.53	.128	-0.01, 0.01
PS	AT (direct effect)	-0.18	0.18	-0.98	.326	-0.53, -0.18
	AT (total effect)	-0.46	0.11	-4.20	< .001**	-0.68, -0.25
	Score	0.01	< 0.01	2.05	.041*	0.01, 0.01
PTF	AT (direct effect)	-0.10	0.13	-0.76	.446	-0.35, 0.15
	AT (total effect)	-0.60	0.08	-7.51	< .001**	-0.76, -0.44
	Score	0.01	< 0.01	4.96	< .001**	0.01, 0.01
OA	AT (direct effect)	-0.22	0.15	-1.51	.133	-0.51, 0.07
	AT (total effect)	-0.42	0.09	-4.56	< .001**	-0.60, -0.24
	Score	< 0.01	< 0.01	1.70	.090	-0.01, 0.01
Score	AT	-50.89	2.09	-24.37	< .001**	-55.00, -46.78

Note: Two-tailed p -values; * $p < .05$. ** $p < .01$. Unstandardized regression coefficients reported. Bootstrap sample size 5000. Abbreviations: Assessment type (AT), Perceived Predictive Validity (PPV), Process Satisfaction (PS), Perceived Test Fairness (PTF), Organizational Attractiveness (OA), Test Performance (Score).

perceived predictive validity (indirect effect $ab = -0.188$, $SE = 0.122$, 95%-CI [-0.426, 0.058] or on organizational attractiveness, indirect effect $ab = -0.199$, $SE = 0.116$, 95%-CI [-0.430, 0.026]).

There were no significant interactions between the relationship of assessment type and any of the applicant reaction variables and age. The relationship between assessment type and perceived predictive validity was moderated by gender, in that men scored higher than women, and more so when they were in the traditional vs. gamified condition. Age was negatively related to task performance, and more so in the gamified condition. Detailed results can be seen in Appendix D.

Discussion

The purpose of this study was to broaden previous research about applicant reactions and attitudes towards gamification onto other assessments and samples, by comparing a gamified version of the WCST with the traditional counterpart. It was hypothesized that applicants view the gamified version more positively in terms of predictive validity and process satisfaction, and in turn test fairness and organizational attractiveness. Moreover, test performance was believed to moderate these relationships.

Contrary to the hypotheses put forward, I discovered that statistically significant differences for the levels of perceived predictive validity, process satisfaction and perceived test fairness between groups in favor of the non-gamified assessment method. Furthermore, there was a difference in participants perceptions of organizational attractiveness. The non-gamified cognitive assessment was seen as better suited to predict future job performance and as fairer compared to the GRA. Participants taking the non-gamified assessment also expressed greater satisfaction with the testing process and found it to be more enjoyable than the gamified cognitive assessment. Moreover, they rated the fictional organization as a more desirable place to work than respondents completing the GRA.

These findings are contrary to previous research where GRA usually elicit favorable responses from candidates (e.g., al-Qallawi & Raghavan, 2022; Georgiou & Lievens, 2022) and are generally more highly regarded than traditional selection techniques (e.g., Georgiou & Nikolaou, 2020; Harman & Brown, 2022). Hommel et al.'s (2022) participants even had higher levels of test acceptance with regards to the GSST than the WCST. However, a majority of their sample were German psychology students, who may have previously been exposed to the WCST in their lectures. As the GSST operates on the same principles but with additional gamified elements, they likely had an easier time understanding what they were asked to do and thus may have had more self-efficacy and fun while completing the assessment. Moreover, to be admitted to a psychology program at a German university, applicants typically need at least a GPA between 1.1 to 1.5 (GPAs range from 1.0 to 6.0 with 1.0 being the best, and 4.0 being the passing grade) (*NC Psychologie Wintersemester 2020/2021*, 2024). Therefore, Hommel et al.'s (2022) participants may have been more intelligent than the average population, increasing their performance and likely their rating of the GSST. On the other hand, this study's sample was more heterogenous and possibly more challenged in understanding the GSST principles. This is underlined by private messages of participants recruited via personal contacts, stating that they "sorted according to what makes sense in terms of content, but it was always incorrect", "didn't understand the test" or "the judgements seemed to be completely independent from my answers". One participant, a marketing manager like the character in the GSST, even commented "Apparently, I'm not good at my job". The fact that individuals from the Prolific sample were more likely to discontinue in the gamified condition may also serve as an indicator that participants struggled to comprehend the GSST's instructions and what they should be doing. Taken together, this might explain the less favourable ratings of the GSST in the current study.

Additionally, the findings showed that the association between test fairness and the assessment method was partially mediated by process satisfaction and perceived predictive

validity. Compared to those who took the GRA, it appears that those who completed the traditional assessment had higher levels of process satisfaction and perceived the assessment to have more predictive validity, which favorably impacted their judgments of test fairness. In other words, individuals' attitudes about an object or procedure are shaped by their beliefs, or subjective assessments about it, which in turn influence their behavioral intentions (Fishbein & Ajzen, 1975). Comparably, the favorable evaluation of the non-gamified assessment's attributes produced a favorable sentiment toward the assessment method.

Furthermore, I discovered that more positive perceptions about a selection assessment translated into more positive perceptions about the company, which is consistent with earlier studies on applicant reactions and personnel selection processes (e.g., Hausknecht et al., 2004). In particular, the results showed that, via process satisfaction and test fairness views, as well as through perceived predictive validity and perceived test fairness, there was a full mediation between the assessment method and organizational attractiveness.

Signaling theory states that people who go through a selection process learn about the qualities of the organization via the signals they receive during the process (Spence, 1973). This study's findings showed that using a game fiction-enhanced assessment method as part of the hiring process reduced applicants' satisfaction with the process and, consequently, their perceptions of fairness, which may have signaled to them that the company may not be as fair and enjoyable to work for as their organizational attractiveness levels were lower than those of applicants using a traditional assessment method. Thus, companies hoping to build their employer brand should consider how applicants perceive their hiring process (Schill et al., 2017).

The supported mediation relationships observed in this study are supported by older personnel selection research (e.g., Hausknecht et al., 2004), but contrary to previous gamification research (Georgiou & Nikolaou, 2020), where the association between assessment method and perceived test fairness via process satisfaction, as well as the

association between assessment method and organizational attractiveness via process satisfaction and then perceived test fairness, were favorable of the GRA. However, Georgiou and Nikolaou (2020) only found mediation paths via process satisfaction but not via perceived predictive validity, which I did in this study. It may be that the effect of gamification on applicant reactions largely depends on the specific GRA, and less on gamification in general. This is supported by Landers and Sanchez's (2022) postulation that the success of gamification and gameful design in personnel selection depends on how well game mechanisms are incorporated into the current system.

Lastly, this study assessed whether test performance moderates the effect of gamification on perceived predictive validity, process satisfaction, perceived test fairness and organizational attractiveness. Contrary to the hypothesis, this was not the case. The difference in the salience of test performance between the conditions did not seem to affect the applicant reaction variables. This can be explained in multiple ways. First, previous studies only found small to moderate positive associations between applicant perceptions and selection test performance (e.g., Chan & Schmitt, 1997; Hausknecht et al., 2004; Oostrom et al., 2012), while correlations with perceived test performance were higher (Hausknecht et al., 2004). Oostrom et al. (2012) discovered that applicants' test performance only affected their post-test reactions through self-assessed test performance. Therefore, test performance might not have been an ideal indicator of the salience of test performance. Instead, future studies should assess whether perceived test performance serves as a better moderator for the effect of assessment method on applicant perceptions.

Secondly, explorative analyses showed that instead of as a moderator, test performance served as a full mediator between assessment type and process satisfaction as well as between assessment type and perceived test fairness. Phrased differently, assessment type only influenced process satisfaction and perceived test fairness through test performance. It is important to note that the GSST is not only the gamified version of the WCST, but also has a

higher difficulty level, i.e. there are differences between the tests beyond the gamified elements (Hommel et al., 2022). The explorative findings demonstrate that the differences in applicant reactions between the experimental groups cannot be fully explained by the assessment method, but test performance also plays a substantial role. Future research is needed to further explore this influence of test performance on applicant reactions. Specifically, studies should assess whether the influence of gamification on applicant reactions remains or even changes its value when test performance is held constant.

Strengths and Practical Implications

By looking at further applicant reaction outcomes that had not previously been examined for the respective assessment procedures, as well as collecting data from new samples and different fictional selection scenarios, this study expanded on the literature on gamification and personnel selection. To the best of my knowledge, this is also the first study exploring the effect of participants' test performance on the relationship between assessment method and applicant reaction outcomes in a GRA context.

The results demonstrate that the use of different game features and assessment methods may have different effects on applicants' views. Therefore, organizations should carefully evaluate new assessments before they are used in real application processes (Chamorro-Premuzic et al., 2017; Landers & Sanchez, 2022). As selection procedure perceptions have an impact on employer branding, employers must maintain a positive image throughout the entire selection process to lower the risk of losing top candidates (Cortini et al., 2019). This is especially relevant in a tight applicant market. Therefore, organizations may need to consider other means than gamification to improve applicant reactions. For example, organizations could ensure that the application and assessment websites are user-friendly and mobile-optimized, as well as simplify and shorten the application forms and procedures (Allen et al., 2007). Additionally, training recruiters to interact with candidates respectfully and professionally could create a more positive impression (Chapman & Webster, 2003).

Lastly, as test performance influenced applicant perceptions, companies may need to find ways to improve the candidate experience for applicants who performed less in the selection process. This may be done by providing timely, customized, and informal rejection letters, and thus positively affecting perceived fairness and re-application intentions (Cortini et al., 2019).

Limitations and Implications for Future Research

There are various limitations with this study that should be resolved in future research. First, this study entailed a hypothetical selection process, which likely involves different motivational traits and results (Landers & Behrend, 2015). It is reasonable to assume that real candidates in an actual selection scenario will be more involved or engaged, leading to higher test scores as well as a different perception of the selection process than participants in a selection simulation. Hausknecht et al. (2004) argue that when assessing job possibilities, applicants consider a variety of elements (such as competing offers or economic concerns), which may minimize the influence of emotions in shaping applicant behavior.

As many of these factors are controllable, research from hypothetical selection settings may therefore reveal stronger relationships, lessening the practical significance of hypothetical research findings. Future research should aim to replicate the current findings in a field study with an authentic selection scenario to increase external validity. As the GSST includes specific storytelling about a marketing agency, future research should also investigate whether applicants in various industries and job roles respond differently to scenarios that are more aligned with their experiences.

Furthermore, an individual's perception of the selection process is significantly influenced by the outcome they obtain, i.e. whether they are hired or not (Bauer et al., 1998). To expand the results of gamification in personnel selection, future research should look at further applicant reaction outcomes, e.g., intentions to accept a job offer, and recommendations intentions.

Moreover, because participants only completed one version of the test, the current research design prevented a direct comparison between the GSST and the WCST. Future studies could use a within-subjects design where each participant undergoes both assessments in a randomized order. This would minimize individual differences, but potential order effects (e.g., fatigue, practice effects) would need to be considered.

However, the GSST incorporates several changes to the original WCST. Therefore, it is not clear if the decreased applicant reactions are due to the game's components, its graphic and work-related game content, higher level of difficulty, or a combination of these. Consequently, extra caution should be used when generalizing our findings because applicants' perceptions may fluctuate depending on the game features and/or assessment procedures used. Previous qualitative research from al-Qallawi and Raghavan (2022) found that reactions to GRA that are negative are caused by certain characteristics of technology, such as the game's design or the existence of bugs, rather than the evaluation itself. Therefore, future studies should carefully evaluate which components show quantifiable effects on acceptance and methodically vary each implementation to gain a greater grasp of the underlying mechanisms. These kinds of investigations may lead to useful advancements in the theoretical foundations of GRAs and gamification.

Conclusion

In conclusion, this study advances our understanding of the impact of gamification on personnel selection by exploring applicant reaction outcomes that had not been evaluated for the respective assessments. A novel perspective on the influence of participants' test performance was introduced. The results emphasize the necessity for careful evaluation before integrating new assessments into real application processes. This is particularly crucial in competitive applicant markets, because organizational reputation significantly affects employer branding and talent acquisition outcomes (Hesse & Mattmüller, 2019, p. 24).

Furthermore, the study underscores the broader implications of test performance on applicant perceptions, highlighting the need for companies to enhance the candidate experience, especially for those who may not excel in the selection process. These findings contribute valuable insights to the field, providing considerations for organizations seeking to optimize their recruitment strategies and maintain a positive employer brand throughout the process.

References

- Allen, D. G., Mahto, R. V. & Otondo, R. F. (2007). Web-based recruitment: Effects of information, organizational brand, and attitudes toward a Web site on applicant attraction. *Journal Of Applied Psychology*, 92(6), 1696–1708.
<https://doi.org/10.1037/0021-9010.92.6.1696>
- Al-Qallawi, S. & Raghavan, M. (2021). A review of online reactions to game-based assessment mobile applications. *International Journal of Selection and Assessment*, 30(1), 14–26. <https://doi.org/10.1111/ijsa.12346>
- Aouam, H., Belmouffeq, B., & Mahil, A. (2023). Applying game-based approaches in personnel selection: a systematic literature review. *Management Dynamics in the Knowledge Economy*, 11(1), 16-29. <https://doi.org/10.2478/mdke-2023-0002>
- Ateş, C., Kaymaz, Ö., Kale, H. E., & Tekindal, M. A. (2019). Comparison of test statistics of nonnormal and unbalanced samples for Multivariate Analysis of Variance in terms of type-I error rates. *Computational and Mathematical Methods in Medicine*, 2019, 1–8.
<https://doi.org/10.1155/2019/2173638>
- Auer, E. M., Mersy, G., Marin, S., Blaik, J. & Landers, R. N. (2021). Using machine learning to model trace behavioral data from a game-based assessment. *International Journal of Selection and Assessment*, 30(1), 82–102. <https://doi.org/10.1111/ijsa.12363>
- Baker, J. (2018, March 7). Care About the People You Don't Hire.
<https://www.linkedin.com/pulse/care-people-you-dont-hire-james-baker/>
- Banse, P. & Buermeyer, U. (Hosts). (2023, May 25). LdN337 Razzia bei Letzter Generation, Ampel streitet über Klimagesetze, Transportgenehmigungen bremsen Windkraftausbau, Neue Ideen für Einwanderung (Prof. Dr. Simon Jäger, Leiter des IZA) [Audio podcast episode]. In *Lage der Nation*.
<https://lagedernation.org/podcast/ldn337-razzia-bei-letzter-generation-ampel-streitet->

[ueber-klimagesetze-transportgenehmigungen-bremsen-windkraftausbau-neue-ideen-fuer-einwanderung-prof-simon-jaeger-leiter-des-iza/](https://doi.org/10.1016/j.jbusres.2019.12.041)

- Bauer, T. N., Maertz, C. P., Dolen, M. R., & Campion, M. A. (1998). Longitudinal assessment of applicant reactions to employment testing and test outcome feedback. *Journal of Applied Psychology, 83*, 892–903. <https://doi.org/10.1037/0021-9010.83.6.892>
- 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>
- Bauer, T. N., Truxillo, D. M., Tucker, J. S., Weathers, V. M., Bertolino, M., Erdogan, B. & Campion, M. A. (2006). Selection in the Information Age: The impact of privacy concerns and computer experience on applicant reactions. *Journal of Management, 32*(5), 601–621. <https://doi.org/10.1177/0149206306289829>
- Barsky, A., Kaplan, S. A. & Beal, D. J. (2010). Just feelings? The role of affect in the formation of organizational fairness judgments. *Journal of Management, 37*(1), 248–279. <https://doi.org/10.1177/0149206310376325>
- Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7th, completely revised and expanded edition). Springer textbook. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg.
- Brislin, R.W. (1980) Translation and content analysis of oral and written material. In: Triandis, H.C. and Berry, J. W., Eds., *Handbook of cross-cultural psychology: Methodology*. Boston: Allyn and Bacon, pp. 389-444.
- Buil, I., Catalán, S. & Martínez, E. P. (2020). Understanding applicants' reactions to gamified recruitment. *Journal of Business Research, 110*, 41–50. <https://doi.org/10.1016/j.jbusres.2019.12.041>

- Cañas, J. J., Quesada, J., Antolí, A., & Fajardo, I. (2003). Cognitive flexibility and adaptability to environmental changes in dynamic complex problem-solving tasks. *Ergonomics*, *46*, 482–501. <https://doi.org/10.1080/0014013031000061640>
- Chamorro-Premuzic, T., Akhtar, R., Winsborough, D., and Sherman, R. A. (2017). The datafication of talent: how technology is advancing the science of human potential at work. *Current Opinion in Behavioral Sciences*, *18*, 13–16. <https://doi.org/10.1016/j.cobeha.2017.04.007>
- Chan, D. & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, *82*(1), 143–159. <https://doi.org/10.1037/0021-9010.82.1.143>
- Chan, D. & Schmitt, N. (2004). An Agenda for Future Research on Applicant Reactions to selection Procedures: A Construct-Oriented Approach. *International Journal of Selection and Assessment*, *12*(1–2), 9–23. <https://doi.org/10.1111/j.0965-075x.2004.00260.x>
- Chan, D., Schmitt, N., Sacco, J. M., & DeShon, R. P. (1998). Understanding pretest and posttest reactions to cognitive ability and personality tests. *Journal of Applied Psychology*, *83*(3), 471. <https://doi.org/10.1037/0021-9010.83.3.471>
- Collmus, A. B., Armstrong, M. B., & Landers, R. N. (2016). Game-thinking within social media to recruit and select job candidates. In R. N. Landers, & G. B. Schmidt (Eds.), *Social media in employee selection and recruitment: Theory, practice, and current challenges* (pp. 103-124). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-29989-1_6
- Collmus, A. B. & Landers, R. N. (2019). Game-Framing to improve applicant perceptions of cognitive assessments. *Journal of Personnel Psychology*, *18*(3), 157–162. <https://doi.org/10.1027/1866-5888/a000227>

Cortini, M., Galanti, T. & Barattucci, M. (2019). The Effect of Different Rejection Letters on Applicants' Reactions. *Behavioral Sciences*, 9(10), 102.

<https://doi.org/10.3390/bs9100102>

Davidson, R., & MacKinnon, J. G. (1993). *Estimation and Inference in Econometrics*. Oxford University Press.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. New York: McGraw-Hill.

Georgiou, K. (2021). Can explanations improve applicant reactions towards gamified assessment methods? *International Journal of Selection and Assessment*, 29(2), 253–268. <https://doi.org/10.1111/ijsa.12329>

Georgiou, K., Gouras, A. & Nikolaou, I. (2019). Gamification in Employee Selection: The development of a gamified assessment. *International Journal of Selection and Assessment*, 27(2), 91–103. <https://doi.org/10.1111/ijsa.12240>

Georgiou, K. & Lievens, F. (2022). Gamifying an assessment method: What signals are organizations sending to applicants? *Journal of Managerial Psychology*, 37(6), 559–574. <https://doi.org/10.1108/jmp-12-2020-0653>

Georgiou, K. & Nikolaou, I. (2020). Are applicants in favor of traditional or gamified assessment methods? Exploring applicant reactions towards a gamified selection method. *Computers in Human Behavior*, 109, 106356.

<https://doi.org/10.1016/j.chb.2020.106356>

Gilliland, S. W. (1993). The Perceived Fairness of Selection Systems: An Organizational Justice Perspective. *Academy of Management Review*, 18(4), 694–734.

<https://doi.org/10.5465/amr.1993.9402210155>

Gkorezis, P., Georgiou, K., Nikolaou, I. & Kyriazati, A. (2020). Gamified or traditional situational judgement test? a moderated mediation model of recommendation intentions via organizational attractiveness. *European Journal of Work and*

Organizational Psychology, 30(2), 240–250.

<https://doi.org/10.1080/1359432x.2020.1746827>

Greenberg, J. (1993). The social side of fairness: Interpersonal and informational classes of organizational justice. In R. Cropanzano (Ed.) *Justice in the workplace* (pp. 79-103).

Hillsdale, NJ: Erlbaum. <http://psycnet.apa.org/psycinfo/1993-97521-004>

Harman, J. L. & Brown, K. (2022). Illustrating a narrative: a test of game elements in game-like personality assessment. *International Journal of Selection and Assessment*, 30(1), 157–166. <https://doi.org/10.1111/ijsa.12374>

Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, 57(3), 639-683. <https://doi.org/10.1111/j.1744-6570.2004.00003.x>.

Hayes, A. F. (2018). *Introduction to Mediation, Moderation, and Conditional Process Analysis, Second Edition (Methodology in the Social Sciences)* (2nd ed.). Guilford Press.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test manual: Revised and expanded*. Lutz, FL: Psychological Assessment Resources.

Herzog, M. H. a., Francis, G., & Clarke, A. (2019). *Understanding statistics and experimental design: How to not lie with statistics. Learning materials in biosciences*. Cham, Switzerland: Springer.

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>

Highhouse, S., Thornbury, E. E. & Little, I. S. (2007). Social-identity functions of attraction to organizations. *Organizational Behavior and Human Decision Processes*, 103(1), 134–146. <https://doi.org/10.1016/j.obhdp.2006.01.001>

- Hommel, B. E., Ruppel, R., & Zacher, H. (2022). Assessment of cognitive flexibility in personnel selection: Validity and acceptance of a gamified version of the Wisconsin Card Sorting Test. *International Journal of Selection and Assessment*, 30, 126–144. <https://doi.org/10.1111/ijsa.12362>
- Küpper, D. M., Klein, K. & Völckner, F. (2021). Gamifying Employer Branding: An integrating framework and research propositions for a new HRM approach in the digitized economy. *Human Resource Management Review*, 31(1), 100686. <https://doi.org/10.1016/j.hrmr.2019.04.002>
- Kluger, A. N., & Rothstein, H. R. (1993). The influence of selection test type on applicant reactions to employment testing. *Journal of Business and Psychology*, 8(1), 3–25. <https://doi-org.proxy-ub.rug.nl/10.1007/BF02230391>
- Landers, R. N., Armstrong, M. B., Collmus, A. B., Mujcic, S. & Blaik, J. (2022). Theory-driven game-based assessment of general cognitive ability: design theory, measurement, prediction of performance, and test fairness. *Journal of Applied Psychology*, 107(10), 1655–1677. <https://doi.org/10.1037/ap10000954>
- Landers, R. N., & Behrend, T. S. (2015). An inconvenient truth: Arbitrary distinctions between organizational, Mechanical Turk, and other convenience samples. *Industrial and Organizational Psychology*, 8, 142–164. <https://doi.org/10.1017/iop.2015.13>
- Landers, R. N. & Sanchez, D. R. (2022). Game-based, gamified, and gamefully designed assessments for employee selection: definitions, distinctions, design, and validation. *International Journal of Selection and Assessment*, 30(1), 1–13. <https://doi.org/10.1111/ijsa.12376>
- Lievens, F. & Highhouse, S. (2003). The relation of instrumental and symbolic attributes to a company's attractiveness as an employer. *Personnel Psychology*, 56(1), 75–102. <https://doi.org/10.1111/j.1744-6570.2003.tb00144.x>

- Macan, T.H., Avedon, M.J., Paese, M., & Smith, D.E. (1994). The effects of applicants' reactions to cognitive ability tests and an assessment center. *Personnel Psychology*, 47, 715-738. <https://doi.org/10.1111/j.1744-6570.1994.tb01573.x>
- Miller, D. T. (1978). What constitutes a self-serving attributional bias? A reply to Bradley. *Journal of Personality and Social Psychology*, 36(11), 1221–1223. <https://doi.org/10.1037/0022-3514.36.11.1221>
- Murphy, K. R. (1986). When your top choice turns you down: Effect of rejected offers on the utility of selection tests. *Psychological Bulletin*, 99(1), 133–138. <https://doi.org/10.1037/0033-2909.99.1.133>
- NC Psychologie Wintersemester 2020/2021. (2024, January 21). <https://www.nc-klage-psychologie.de/numeris-clausus/nc-psychologie.html>
- Nikolaou, I. (2021). What is the role of technology in recruitment and selection? *Spanish Journal of Psychology*, 24. <https://doi.org/10.1017/sjp.2021.6>
- Nikolaou, I., Georgiou, K., Bauer, T. N., & Truxillo, D. M. (2019a). Technology and applicant reactions. In R. N. Landers (Ed.), *Cambridge handbook of technology and employee behavior*. Cambridge: Cambridge University Press.
- Nikolaou, I., Georgiou, K. & Kotsasarlidou, V. (2019). Exploring the relationship of a gamified assessment with performance. *Spanish Journal of Psychology*, 22. <https://doi.org/10.1017/sjp.2019.5>
- OECD (2016). *Getting skills right: Assessing and anticipating changing skill needs*. Paris: OECD Publishing.
- Oostrom, J. K., Bos-Broekema, L. , Serlie, A. W., Born, M. Ph. & Van der Molen, H. T. (2012). A Field Study of Pretest and Posttest Reactions to a Paper-and-Pencil and a Computerized In-Basket Exercise, *Human Performance*, 25(2), 95-113, <https://doi.org/10.1080/08959285.2012.658928>

- Palan, S. & Schitter, C. (2017). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27.
<https://doi.org/10.1016/j.jbef.2017.12.004>
- Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153-163. <https://doi.org/10.1016/j.jesp.2017.01.006>
- Ployhart, R. E. & Harold, C. M. (2004). The Applicant Attribution-Reaction Theory (AART): an integrative theory of applicant attributional processing. *International Journal of Selection and Assessment*, *12*(1–2), 84–98. <https://doi.org/10.1111/j.0965-075x.2004.00266.x>
- Ployhart, R. E. & Ryan, A. M. (1997). Toward an explanation of applicant reactions: an examination of organizational justice and attribution frameworks. *Organizational Behavior and Human Decision Processes*, *72*(3), 308–335.
<https://doi.org/10.1006/obhd.1997.2742>
- Ployhart, R.E., & Ryan, A.M. (1998). Applicants' reactions to the fairness of selection procedures: The effects of positive rule violations and time of measurement. *Journal of Applied Psychology*, *83*, 3-16. <https://doi.org/10.1037/0021-9010.83.1.3>
- PricewaterhouseCoopers (2012). Millennials at work - Reshaping the workplace. Retrieved August 8, 2023 from <https://store.pwc.de/de/publications/millennials-at-work-reshaping-the-workplace>
- Ramos-Villagrasa, P. J., Del Río, E. F. & Castro, Á. (2022). Game-related Assessments for Personnel Selection: A Systematic review. *Frontiers in Psychology*, *13*.
<https://doi.org/10.3389/fpsyg.2022.952002>
- Ryan, A. M. & Ployhart, R. E. (2000). Applicants' Perceptions of Selection Procedures and Decisions: A Critical Review and Agenda for the Future. *Journal of Management*, *26*(3), 565–606. <https://doi.org/10.1177/014920630002600308>

- Ryan, A. M., & Ployhart, R. E. (2014). A century of selection. *Annual Review of Psychology*, 65(1), 693–717. <https://doi.org/10.1146/annurev-psych-010213-115134>
- Salgado, J.F., Moscoso, S., García-Izquierdo, A.L., Anderson, N.R. (2017). Inclusive and Discrimination-Free Personnel Selection. In: Arenas, A., Di Marco, D., Munduate, L., Euwema, M. (Eds.) *Shaping Inclusive Workplaces Through Social Dialogue. Industrial Relations & Conflict Management*. Springer, Cham. https://doi.org/10.1007/978-3-319-66393-7_7
- Sanchez, D. R., Weiner, E. & Van Zelderen, A. (2022). Virtual Reality Assessments (VRAS): Exploring the reliability and validity of evaluations in VR. *International Journal of Selection and Assessment*, 30(1), 103–125. <https://doi.org/10.1111/ijsa.12369>
- Schill, D., Kracklauer, A. H. & Rasche, C. (2017). The Quality of the Candidate Journey as an Important Aspect of the Employer Brand. AICSSH & AICLEF 2017 (Cambridge) Conference Proceedings. <https://publications.hs-neu-ulm.de/1381/>
- Schmit, M. J. & Ryan, A. M. (1997). Applicant withdrawal: The role of test-taking attitudes and racial differences. *Personnel Psychology*, 50(4), 855–876. <https://doi.org/10.1111/j.1744-6570.1997.tb01485.x>
- Sears, G. J., Zhang, H., Wiesner, W. H., Hackett, R. D. & Yuan, Y. (2013). A comparative assessment of videoconference and face-to-face employment interviews. *Management Decision*. <https://doi.org/10.1108/md-09-2012-0642>
- Smither, J. W., Reilly, R. R., Millsap, R. E., At, T, K. P., & 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>
- Spardel, L. (2023, 22 May). Fachkräftemangel: Welche Ursachen und Lösungen gibt es? Recrutee. <https://recrutee.com/de-artikel/fachkraeftemangel-ursachen-loesungen>
- Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics*, 87(3), 355. <https://doi.org/10.2307/1882010>

- Steiner, K. (2017). Bad candidate experience cost Virgin Media \$5M annually – Here is how they turned that around. Retrieved from:
<https://business.linkedin.com/talentsolutions/blog/candidate-experience/2017/bad-candidate-experience-cost-virginmedia-5m-annually-and-how-they-turned-that-around>
- Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44*, 24–31.
<https://doi.org/10.1177/0098628316677643>
- Stone, E. R. (2010). t Test, Independent Samples. In N. J. Salkind (Ed.), *Encyclopedia of research design* (pp. 1551–1556). Los Angeles: SAGE.
- Sylva, H. & Mol, S. T. (2009). E-Recruitment: A study into applicant perceptions of an online application system. *International Journal of Selection and Assessment*, *17*(3), 311–323. <https://doi.org/10.1111/j.1468-2389.2009.00473>
- Truxillo, D. M., Bauer, T. N., McCarthy, J. M., Anderson, N. & Ahmed, S. M. (2016). Applicant Perspectives on Employee selection Systems. In D. Ones, N. Anderson, C. Viswesvaran, & H. Sinangil (Eds.), *Handbook of industrial, work and organizational psychology (2nd ed.)*. London: Sage. <https://doi.org/10.4135/9781473914940.n19>
- Wiernik, B. M., Raghavan, M., Caretta, T. R. & Coovert, M. D. (2022). Developing and validating a serious game-based assessment for cyber occupations in the US Air Force. *International Journal of Selection and Assessment*, *30*(1), 27–47.
<https://doi.org/10.1111/ijasa.12378>
- Woods, S. A., Ahmed, S. & Nikolaou, I., Coste, A. C., Neil, A. (2020) Personnel selection in the digital age: A review of validity and applicant reactions, and future research challenges. *European Journal of Work and Organizational Psychology*, *29*(1), 64-77.
<https://doi.org/10.1080/1359432X.2019.1681401>
- Wu, F. Y., Mulfinger, E., Alexander, L., Sinclair, A. L., McCloy, R. A. & Oswald, F. L. (2021). Individual differences at play: an investigation into measuring big five

personality facets with game-based assessments. *International Journal of Selection and Assessment*, 30(1), 62–81. <https://doi.org/10.1111/ijsa.12360>

Appendix A

Fictional application scenario

English version

Instructions: *Below you will read about a job application scenario. Please take the perspective of someone who is actually in this situation.*

Imagine that you are seeking a new opportunity to advance your career in the field of marketing. You come across an exciting job opening for the position of a Marketing Director at a prestigious and innovative company. The company, Horizon Marketing Solutions, is known for its cutting-edge strategies, creative campaigns, and strong market presence.

Feeling intrigued by the position, you carefully review the job description and qualifications. You realize that this opportunity aligns perfectly with your skills, experience, and ambitions. The thought of being at the forefront of marketing initiatives, leading a team of talented professionals, and making a significant impact in the industry fills you with enthusiasm.

Encouraged by your strong belief in your abilities and the potential match with the company's vision, you decide to take the next step and submit your application. Soon after, you receive an email expressing interest from the company's HR department, inviting you to participate in an online cognitive ability assessment as the first part of the application process:

“Dear applicant,

Thank you for your interest in the Marketing Director position at Horizon Marketing Solutions. We are excited to invite you to participate in our selection process. As the first part of our comprehensive assessment, we ask you to complete a Cognitive Ability Test to evaluate your cognitive flexibility, a critical skill for the Marketing Director role. This online test will assess your critical thinking, problem-solving abilities, and adaptability in dynamic scenarios. Your performance in this assessment will help us assess your potential to drive innovation, formulate effective strategies, and make decisions as a Marketing Director.

Thank you for your participation and commitment to this selection process. We appreciate your time and effort in demonstrating your cognitive abilities for the Marketing Director position at Horizon Marketing Solutions.”

Now, with this scenario in mind, please continue to the next page and immerse yourself in the online assessment. **The assessment can only be completed on a laptop/computer. Please maximize your browser window.**

German version (shown to participants)

Instructions: *Im Folgenden lesen Sie über ein Bewerbungsszenario. Bitte nehmen Sie die Perspektive von jemandem ein, der/die sich tatsächlich in dieser Situation befindet.*

Stellen Sie sich vor, Sie sind auf der Suche nach einer neuen Möglichkeit, Ihre Karriere im Marketing voranzutreiben. Sie stoßen auf eine spannende Stellenausschreibung für die Position eines Marketing Directors bei einem angesehenen und innovativen Unternehmen. Das Unternehmen, Horizon Marketing Solutions, ist für seine innovativen Strategien, kreativen Kampagnen und seine starke Marktpräsenz bekannt.

Sie fühlen sich von der Stelle angesprochen und sehen sich die Stellenbeschreibung und die Qualifikationen genau an. Sie stellen fest, dass diese Gelegenheit perfekt zu Ihren Fähigkeiten, Ihrer Erfahrung und Ihren Ambitionen passt. Der Gedanke, an der Spitze von Marketinginitiativen zu stehen, ein Team von talentierten Fachleuten zu leiten und einen bedeutenden Einfluss auf die Branche auszuüben, begeistert Sie.

Da Sie fest von Ihren Fähigkeiten überzeugt sind und die Vision des Unternehmens auf Sie zutrifft, beschließen Sie, sich zu bewerben. Kurz darauf erhalten Sie eine E-Mail von der Personalabteilung des Unternehmens, in der Sie gebeten werden, als ersten Teil des Bewerbungsverfahrens an einem Online-Assessment zur Beurteilung Ihrer kognitiven Fähigkeiten teilzunehmen:

"Liebe:r Beweber:in,

vielen Dank für dein Interesse an der Stelle des Marketing Directors bei Horizon Marketing Solutions. Wir freuen uns, dich einladen zu können, an unserem Auswahlverfahren teilzunehmen. Als ersten Teil unseres umfassenden Assessments bitten wir dich, einen kognitiven Fähigkeitstest zu absolvieren, um deine kognitive Flexibilität zu testen, eine entscheidende Fähigkeit für die Position des Marketing Directors. Mit diesem Online-Test werden dein kritisches Denken, Problemlösungsfähigkeiten und Anpassungsfähigkeit in einer dynamischen Umgebung bewertet. Deine Leistung in diesem Test wird uns helfen, dein Potenzial einzuschätzen, Innovationen voranzutreiben, effektive Strategien zu formulieren und Entscheidungen als Marketing Director zu treffen.

Wir danken dir für deine Teilnahme und dein Engagement in diesem Auswahlverfahren. Wir wissen es zu schätzen, dass du dir die Zeit nimmst, deine Fähigkeiten für die Stelle des Marketing Directors bei Horizon Marketing Solutions zu zeigen."

Bitte fahren Sie nun, mit dem Szenario im Hinterkopf, auf der nächsten Seite fort und tauchen Sie in das Online-Assessment ein. **Das Assessment kann ausschließlich an einem Laptop/Computer durchgeführt werden. Bitte maximieren Sie ihr Browser-Fenster.**

Appendix B

Questionnaire

Instructions in English

“On this page, you will find a series of statements that relate to your experience with the technology-based assessment you just completed. Please read each statement carefully and indicate the extent to which you agree.”

Instructions in German

„Auf dieser Seite finden Sie eine Reihe von Aussagen, die sich auf Ihre Erfahrungen mit dem technologiegestützten Assessment beziehen, das sie gerade absolviert haben. Bitte lesen Sie jede Aussage sorgfältig durch und geben Sie an, inwieweit Sie zustimmen.“

Table 6

Questionnaire with Translations

Scale	Original item	German translation
	I think this test is fair.	Ich denke, dieser Test ist fair.
	Most people would say that this test is fair.	Die meisten Leute würden sagen, dass dieser Test fair ist.
Perceived test fairness	I believe that this test can predict whether I will be a successful employee.	Ich glaube, dieser Test kann vorhersagen, ob ich ein:e erfolgreiche:r Mitarbeiter:in sein werde.
	I can see the connection between this test and performance on the job.	Ich sehe den Zusammenhang zwischen diesem Test und der Arbeitsleistung.
Perceived predictive validity	I am confident that the test can predict how well an applicant will perform on the job.	Ich bin zuversichtlich, dass der Test vorhersagen kann, wie gut ein:e Bewerber:in bei der Arbeit sein wird.
	The employer can tell a lot about the applicant's ability to do the job based on the results of the test.	Der/Die Arbeitgeber:in kann anhand der Testergebnisse viel über die Fähigkeit des/der Bewerber:in, die Arbeit zu erledigen, sagen.
	Failing to perform well on the test indicates that the applicant cannot	Ein schlechtes Abschneiden beim Test deutet darauf hin, dass der/die Bewerber:in bei der

	perform well on the job.	Arbeit nicht gut abschneiden kann.
Process satisfaction	Overall, I was satisfied with this application process.	Insgesamt war ich mit diesem Bewerbungsverfahren zufrieden.
	For me, this company would be a good place to work.	Für mich wäre dieses Unternehmen ein guter Ort zum Arbeiten.
	I would not be interested in this company except as a last resort.	Ich wäre nicht an diesem Unternehmen interessiert, außer als letzter Ausweg.
Organizational attractiveness	This company is attractive to me as a place for employment.	Dieses Unternehmen ist für mich als Beschäftigungsort attraktiv.
	I am interested in learning more about this company.	Ich bin daran interessiert, mehr über dieses Unternehmen zu erfahren.
	A job at this company is very appealing to me.	Ein Arbeitsplatz in diesem Unternehmen ist für mich sehr ansprechend.

Appendix C

Preliminary Analyses

Testing Assumptions for Hypothesis 1

Boxplots were created to inspect possible univariate outliers. Multivariate outliers were assessed using the Mahalanobis distance (cut-off value $\chi^2(df = 3) = 16.266, p = .01$). Two univariate outliers were found in perceived test fairness and six for perceived predictive validity for the gamified condition. In addition, two multivariate outliers were found. Respective participants were not excluded from further analysis, as I did not want to exclude anyone for having strong opinions on an assessment. The normality assumption was visually inspected with histograms and Q-Q-plots as well as statistically tested with the Shapiro-Wilk test ($\alpha = .05$). Shapiro-Wilk tests were significant for both conditions across perceived predictive validity, process satisfaction and test fairness perceptions. As MANOVA is relatively robust against violations of the normal distribution (Finch, 2005), the analysis was continued. Lastly, correlations between dependent variables were assessed to rule out multicollinearity. Since correlations between the respective variables were low ($r < .90$), multicollinearity was ruled out as a confounding factor in the analysis.

For perceived test fairness and process satisfaction, there was homogeneity of the error variances ($p > .05$), but not for perceived predictive validity ($p = .007$), as assessed by Levene's tests. Moreover, there was no homogeneity of covariances, as assessed by Box's test ($p = .034$). As SPSS calculates multiple statistics which are relatively robust against heterogeneity of error variances and covariances (Ateş et al., 2019), the analysis was continued.

Testing Assumptions for Hypothesis 3

As the sample was sufficiently large ($n > 30$ for each of the two groups), normal distribution was not tested, since according to the central limit theorem, the sampling distribution will be approximately normally distributed (Bortz & Schuster, 2010; Herzog et al.

, 2019, p. 56; Stone, 2010, p. 1554). Outliers were visually inspected using boxplots, and there were no outliers in the data. Homogeneity of variances was checked via the Levene-test, which was not significant, $F = 0.087, p = .768$.

Testing Assumptions for Hypotheses 2, 4, and 5

The linearity assumptions were assessed by visual inspection of the scatterplots after LOESS smoothing, which revealed that all the dependent variables had a roughly linear relationship.

Appendix D

Detailed Results from Exploratory Analyses

Table 7

Results of Moderation Analyses with Age as Moderator

Y	Predictors	B	SE	t	p (two-tailed)	95% CI
PPV	AT	0.14	0.27	0.51	.612	-0.39, 0.66
	Age	0.02	0.02	1.51	.132	-0.01, 0.05
	AT * Age	-0.01	0.01	-1.42	.158	-0.03, 0.01
Overall model summary: $F(3, 386) = 2.71, p = .045, R^2 = .023$						
PS	AT	-0.35	0.33	-1.06	.292	-1.01, 0.303
	Age	0.01	0.02	0.54	.589	-0.03, 0.04
	AT * Age	< -0.01	0.01	-0.38	.704	-0.03, 0.02
Overall model summary: $F(3, 386) = 5.97, p < .001, R^2 = .045$						
PTF	AT	-0.50	0.25	-2.02	.044*	-0.98, -0.01
	Age	< 0.01	0.01	0.18	.859	-0.02, 0.03
	AT * Age	< -0.01	0.01	-0.43	.671	-0.02, 0.01
Overall model summary: $F(3, 386) = 19.66, p < .001, R^2 = .130$						
OA	AT	-0.56	0.28	-1.99	.048*	-1.12, -0.01
	Age	-0.01	0.01	-0.69	.490	-0.04, 0.02
	AT * Age	< 0.01	0.01	0.52	.607	-0.01, 0.02
Overall model summary: $F(3, 386) = 7.59, p > .001, R^2 = .054$						
Score	AT	-37.68	5.56	-6.77	< .001**	-48.62, -26.74
	Age	0.20	0.24	0.84	.401	-0.27, 0.67
	AT * Age	-0.40	0.15	-2.62	.009*	-0.71, -0.10
Overall model summary: $F(3, 386) = 250.48, p > .001, R^2 = .642$						

Note: Two-tailed p-values; * $p < .05$. ** $p < .01$. Unstandardized regression coefficients

reported. Bootstrap sample size 5000. Abbreviations: Assessment type (AT), Task

Performance (Score), Perceived Predictive Validity (PPV), Process Satisfaction (PS),

Perceived Test Fairness (PTF), Organizational Attractiveness (OA)

Table 8

Results of Moderation Analyses with Age as Moderator

Y	Predictors	B	SE	t	p (two-tailed)	95% CI
PPV	AT	0.34	0.26	1.30	.196	-0.17, 0.86
	Gender	0.74	0.29	2.54	.011*	0.17, 1.30

	AT * Gender	-0.37	0.18	-2.05	.041*	-0.72, -0.02
Overall model summary: $F(3, 380) = 4.035, p = .008, R^2 = .037$						
PS	AT	-0.28	0.35	-0.81	.417	-0.96, 0.40
	Gender	0.31	0.36	0.87	.383	-0.39, 1.02
	AT * Gender	-0.10	0.23	-0.46	.645	-0.55, 0.34
Overall model summary: $F(3, 380) = 6.16, p < .004, R^2 = .045$						
PTF	AT	-0.25	0.25	-1.00	.316	-0.73, 0.24
	Gender	0.43	0.25	1.74	.083	-0.06, 0.92
	AT * Gender	-0.23	0.16	-1.42	.156	-0.55, 0.09
Overall model summary: $F(3, 380) = 20.69, p < .001, R^2 = .135$						
OA	AT	-0.34	0.30	-1.15	.252	-0.92, 0.24
	Gender	0.13	0.29	0.64	.669	-0.45, 0.70
	AT * Gender	-0.05	0.19	-0.24	.809	-0.42, 0.33
Overall model summary: $F(3, 380) = 6.78, p = .002, R^2 = .050$						
Score	AT	-57.77	6.65	-8.68	< .001**	-70.85, -44.68
	Gender	-5.15	5.79	-0.89	.374	-16.54, 6.23
	AT * Gender	4.47	4.28	1.11	.269	-3.68, 13.16
Overall model summary: $F(3, 380) = 193.09, p > .001, R^2 = .614$						

Note: Two-tailed p -values; * $p < .05$. ** $p < .01$. Unstandardized regression coefficients

reported. Bootstrap sample size 5000. Only participants indicating they were female (1) or male (2) were included in the analyses. Abbreviations: Assessment type (AT), Task Performance (Score), Perceived Predictive Validity (PPV), Process Satisfaction (PS), Perceived Test Fairness (PTF), Organizational Attractiveness (OA).