8-2024

It's just a game! Effects of fantasy in a storified test on applicant reactions

Marie L. OHLMS

Klaus G. MELCHERS

Filip LIEVENS
Singapore Management University, filiplievens@smu.edu.sg

Received: 13 March 2024 | Accepted: 15 August 2024

DOI: 10.1111/apps.12569

ORIGINAL ARTICLE

APPLIED PSYCHOLOGY IAAP

It's just a game! Effects of fantasy in a storified test on applicant reactions

Marie L. Ohlms[1,2] | **Klaus G. Melchers**[1] | **Filip Lievens**[3]

[1]Institut für Psychologie und Pädagogik, Universität Ulm, Ulm, Germany

[2]Institut für Psychologie, Albert-Ludwigs-Universität Freiburg, Freiburg, Germany

[3]Lee Kong Chian School of Business, Singapore Management University, Singapore, Singapore

Correspondence
Marie L. Ohlms, Institut für Psychologie, Albert-Ludwigs-Universität Freiburg, Engelbergerstraße 41, D-79085 Freiburg, Germany.
Email: marie.ohlms@psychologie.uni-freiburg.de

Funding information
This work was partially supported by a doctoral scholarship of the Studienstiftung des deutschen Volkes for Marie Luise Ohlms.

Abstract

There is increasing attention to storification of assessments (i.e., embedding a storyline into a non-storified assessment) in research and practice and to gamified and game-based assessment in general. However, there is a surprising lack of agreement and of recommendations regarding what level of fantasy of the storyline one should choose for the storification from the perspective of applicant reactions. A distinction is typically made between fantasy (e.g., fighting aliens) and realistic (e.g., workday simulations) storylines, with both choices having their advantages and disadvantages. In this study, a sample of 195 participants was shown either a storified realistic test, a storified fantasy test, or a non-storified test. Afterwards, they rated various applicant reaction measures. Both storified assessments were rated equally positively on perceived modernity of the organization and enjoyment but the storified realistic test was superior to the storified fantasy test in terms of perceived job-relatedness, procedural fairness, organizational attractiveness, and clarity of work activity. Thus, the level of fantasy of a storyline in a storified assessment plays an important role for applicant reaction variables, whereby the overall pattern of results showed that the storified realistic test was rated most favorably, followed by the non-storified test, and the storified fantasy assessment.

KEYWORDS

applicant reactions, gamification, Germany, job-relatedness, personnel psychology, personnel selection, storification, work and organizational psychology

INTRODUCTION

In recent years, selection procedures using gamified and game-based assessments (GBAs), have gained popularity (Woods et al., 2020). In personnel selection, gamified assessment refers to the introduction of game elements such as levels or storylines into traditional procedures and thus represents a re-design strategy (Landers & Sanchez, 2022). In contrast, GBA as a stand-alone method refers to (computer) games designed to assess applicants' knowledge, skills, abilities, or other characteristics (Landers & Sanchez, 2022). Gamified assessment and GBA are commonly grouped under the umbrella term game-related assessment (GRA).

In the literature (e.g., see Bhatia & Ryan, 2018; Fetzer et al., 2017) as well as in descriptions from GRA providers, improved applicant reactions as compared to their traditional counterparts are described as a major advantage of GRAs. According to proponents of GRAs, the introduction of game elements into assessments may help to increase the enjoyment of the assessment as well as to reduce negative reactions to testing, such as test anxiety (Bhatia & Ryan, 2018; Woods et al., 2020). Additionally, the use of these GRAs may bolster the image of an organization as an employer, because applicants might interpret the innovative features associated with a gamified assessment as characteristics of the organization (Georgiou & Lievens, 2022).

When gamifying a traditional assessment (e.g., a cognitive ability test), one can choose from various game elements. One such gamification approach is storification, which means the embedding of a storyline (also called a narrative or cover story) into a traditional non-gamified assessment to add context and purpose to it (Akkerman et al., 2009; Fetzer et al., 2017). However, a thorny issue that researchers and practitioners frequently face when developing storified assessments is whether to use a fantasy storyline (such as taking the role of a character experiencing adventures on an island, e.g., Georgiou et al., 2019) or a realistic storyline (e.g., the simulation of a fictitious organization one must manage; Melchers & Basch, 2022). The choice between a fantasy and a realistic storyline is important because both storification approaches have their advantages and disadvantages.

First, the level of fantasy of a storyline might affect applicant reactions to gamified assessments. When a fantasy storyline is used that is unrelated to the actual job, the storified assessment may appear similar to a traditional entertainment game. Accordingly, applicants may perceive such a fantasy storified test as enjoyable, might become immersed in the game world, and thus enter a state of flow so that they might even forget about being tested (Bhatia & Ryan, 2018). However, the flip side is that a fantasy storyline might make applicants wonder why they must complete such an assessment that has little or nothing to do with showing their appropriate job skills for the job they are applying for. Conversely, if one chooses a realistic storyline, storification might support the perceived job relatedness of the test and might even be used as a job preview by conveying information about the organization or the job through the storyline. Yet, applicants might consider such a realistic storyline less entertaining and fun than a fantasy storyline. Thus, without evidence on what level of fantasy of a storyline one should

choose from the applicants' point of view, practitioners as well as researchers are left in the dark about the best design of storified tests. Secondly, the choice of a storyline can also have an impact on the effectiveness and efficiency of the recruitment process, because it might influence to what extent applicants perceive the organization to be a modern employer so that they are more likely to stay in the recruitment process and ultimately accept a job offer (Hausknecht et al., 2004).

The present study focused on contrasting these perspectives (i.e., using a fantasy vs. a realistic storyline) as well as comparing storified procedures with a traditional non-gamified test (as a baseline). By doing so, we aimed to gain a better understanding of whether storified assessments lead to better test taker reactions than more traditional procedures and if so, what level of fantasy should be chosen to improve applicant reactions. Accordingly, we compared various applicant reaction variables towards three different versions of a cognitive ability test: a non-storified test, a storified realistic test, and a storified fantasy test. By doing so, we could isolate the effect of the level of fantasy of a storified assessment on applicant reactions.

Our study contributes to both theory and practice of gamified (storified) assessments. Theoretically, based on Gilliland's (1993) fairness model and signaling theory (Spence, 1973), we enhance understanding of how the level of fantasy affects test takers' reactions. For practitioners, the study highlights whether storified assessments should be used from the perspective of applicant reactions and, if so, what level of fantasy should be applied. Therefore, our results may help to guide decisions regarding the choice of the type of fantasy in the development of storified assessments.

ASSESSMENT STORIFICATION

There is a wide range of possible game elements that can be used within a gamified assessment (Bedwell et al., 2012; Fetzer et al., 2017), and many gamified assessments employ several different game elements (see Georgiou & Nikolaou, 2020, for an example). However, to draw conclusions about the impact of specific game elements on applicant reactions, a specific element needs to be considered in isolation so that any differences in reactions compared to more traditional methods can be attributed to a specific re-design process (Lievens & Sackett, 2017). One such game element used in many playful assessments is a storyline (e.g., Georgiou & Nikolaou, 2020; Ohlms et al., 2024a). This gamification approach of integrating a traditional non-gamified test into a storyline is called storification (Akkerman et al., 2009). Hence, the key feature of storification is that the actual test items are framed in a coherent, stringent, and engaging storyline, aiming to capture the attention of test takers and to create immersion (Fetzer et al., 2017; Toda et al., 2019).

Given that not all storylines are the same, assessment developers have wide freedom when designing a cover story. Thus, generally, any scenario is imaginable to link the different parts of a test together. A particularly important decision during the storification process is whether to choose a realistic or a fantasy storyline to connect the different test items with each other. Fantasy describes the creation of a game environment that is detached from the real world and evokes mental associations that do not actually exist (see Malone, 1981; Malone & Lepper, 1987, elsewhere, this is also referred to as [game] fiction, e.g., Sailer & Homner, 2020; Bedwell et al., 2012). For instance, a storified assessment in which the storyline focuses on relieving a country from its curse (e.g., Ohlms et al., 2024a) or having adventures on an island (e.g., Georgiou & Nikolaou, 2020) has no relation to the tasks of an office job (i.e., fantasy

context). Similarly, playing mini-games with a puzzle-like format (Landers et al., 2022) might appear unrelated to most jobs (i.e., low psychological fidelity). In contrast, a simulation of a fictitious organization that applicants have to manage by completing the different test items in a storified assessment demonstrates job-relatedness for the job activities of a manager (i.e., realistic context). Thus, the level of fantasy can differ fundamentally between different storylines.

It is important to note that in addition to the decision about the level of fantasy of a storyline, there are other design decisions in storyline construction that can affect the quality of the resulting assessment and the applicants' reaction to it. For example, a decision must be made about the contingency of the storyline, that is, whether the storyline will always continue in the same way or whether it will be based on an applicant's performance (see also Reddock et al., 2020, for a detailed discussion). In our study, we chose to focus on the level of fantasy, holding other storyline design decisions constant.

RESEARCH ON APPLICANT REACTIONS TO STORIFIED ASSESSMENTS

To the best of our knowledge, few published studies have so far focused on applicant reactions to gamified assessments involving a storyline (e.g., Georgiou & Nikolaou, 2020; Harman & Brown, 2022). In fact, only two published studies in the context of selection and assessment have examined the effects of storification on applicant reactions in isolation. First, Harman and Brown (2022) revised and used a storified test developed by McCord et al. (2019). In their study, a traditional personality test was embedded in a storyline in which the protagonist woke up in a cave with no memory and had to complete various tasks in a fantasy world. Harman and Brown found that participants rated the storified test significantly higher on enjoyment and ease of use and significantly lower on perceived effort than the traditional non-storified personality assessment.

Secondly, Landers and Collmus (2022) examined whether the storification of a traditional openness and conscientiousness measure might improve applicant reactions towards it. To storify the non-storified test, a narrative was used in which participants adopted the role of a college student and experienced a day in this student's everyday life. Landers and Collmus found small positive effects for their storified assessment on perceived predictive validity and fairness but no difference from the traditional assessment in terms of face validity.

Taken together, the limited available evidence on applicant reactions towards storification in personnel selection suggests that storified assessments may potentially improve reactions relative to more traditional instruments (cf. Harman & Brown, 2022; Landers & Collmus, 2022). Yet previous research has not examined the important issue of whether the level of fantasy of the storyline of an assessment impacts reactions towards it.

Although there are no studies in the context of personnel selection on whether a fantasy vs. realistic storyline influences applicant reactions, there is research in the educational field on the relevance of the level of fantasy on crucial outcomes such as learning (e.g., Aldemir et al., 2018; Bai et al., 2022; Parker & Lepper, 1992; Sailer & Homner, 2020). For example, Bai et al. (2022) found that integrating a fantasy storyline (i.e., saving a princess from a dragon) into an online course led to increased peer interaction, learning experience, and learning performance compared to an online course without such a fantasy storyline. Furthermore, a qualitative study by Aldemir et al. (2018) showed that while some participants enjoyed a fantasy

storyline in a gamified course (i.e., the storyline was inspired by the Harry Potter series), other participants stated that the storyline should have some relevance to the real world.

However, it is unclear to what extent the findings from the educational field on the application of fantasy game elements can be generalized to the context of personnel selection. For instance, the populations of interest in educational research are typically children or students, whereas personnel selection research aims to draw conclusions about applicants. Furthermore, while fantasy elements may increase learner engagement and learning outcomes in educational contexts (Sailer & Homner, 2020), fantasy elements may undermine the perceived seriousness of the selection process. Accordingly, job applicants may perceive a fantasy storyline of an assessment as less professional or relevant to the actual job, leading to skepticism about the validity and reliability of the selection test. Thus, to draw sound theoretical and practical conclusions, it is important to examine the effects of a fantasy vs. realistic storyline in a personnel selection context.

THEORETICAL BACKGROUND AND DEVELOPMENT OF HYPOTHESES

Perceived job-relatedness addresses the degree to which an assessment either seems to capture content related to a specific work task or seems to be valid (Gilliland, 1993). The appearance of validity can refer to applicants either perceiving the test content to be related to the content of the work tasks (i.e., content validity) or believing that their test performance is predictive of their later task performance (i.e., criterion-related validity, Gilliland, 1993). As noted above, depending on the specific appearance of a test, the perceived job-relatedness of a storified assessment to the actual tasks of a vacant position may also differ substantially between different storified tests.

Storification of a traditional cognitive ability test by embedding it in an activity-related realistic storyline, such as assisting potential colleagues in solving work problems, might improve applicants' perceived job-relatedness. This is because gamification by means of a realistic storyline may help to increase the relation between the test content and the actual job activity (Gilliland, 1993). In contrast, if a traditional cognitive ability test is embedded in a storyline about fighting aliens, for example, this storyline is completely unrelated to the actual job so that applicants might ask themselves why they must complete such a test in order to be offered a job. Since successfully fighting aliens is not relevant to most occupations, applicants may see poor job-relatedness in such a fantasy storification. Along these lines, we propose the following:

> **Hypothesis 1.** Perceived job-relatedness is higher for a storified realistic test than for a storified fantasy test.[1]

> **Hypothesis 2.** Perceived job-relatedness is lower for a non-storified test than for a storified realistic test, but higher than for a storified fantasy test.

According to Gilliland (1993), perceived job-relatedness is the most influential predictor of perceived general procedural fairness. Additionally, meta-analytic evidence supports the strong positive influence of perceived job-relatedness on applicant fairness perceptions of various selection instruments (Hausknecht et al., 2004). Furthermore, perceived job-relatedness is linked to more favorable perceptions of the organization as well as to a higher willingness to accept a job

offer and to recommend the organization to others (see the meta-analytic results from Chapman et al., 2005; and from Hausknecht et al., 2004). Thus, perceived job-relatedness is among the key factors affecting perceptions of general fairness, organizational attractiveness, and behavioral intentions (Chapman et al., 2005; Hausknecht et al., 2004). In light of these findings, the level of fantasy of the storyline of an assessment test may also affect other variables of Gilliland's fairness model. Furthermore, it is reasonable to assume that a fantasy storification approach, in which test takers do not perceive any association between test tasks and future job activities, might negatively affect attitudes toward the organization as well as behavioral intentions (i.e., acceptance of a job offer, recommendation intention). Furthermore, initial findings on applicant reactions to gamified assessments suggest that the perception of job-relatedness is a more relevant factor for general fairness perceptions and attitudes towards the organization than the perceived enjoyment or modernity of the assessment (Ohlms et al., 2023). Drawing on the extensive findings on the perceived job-relatedness of selection instruments, we hypothesize the following:

> **Hypothesis 3.** (a) Procedural fairness, (b) organizational attractiveness, and (c) behavioral intentions are higher for a storified realistic test than for a storified fantasy test.

> **Hypothesis 4.** (a) Procedural fairness, (b) organizational attractiveness, and (c) behavioral intentions are lower for a non-storified test than for a storified realistic test, but higher than for a storified fantasy test.

Although the focus of the expected benefits of gamification lies in improving applicant reactions, the storification approach may offer further advantages beyond simply increasing test-takers' attitudes. For instance, Ohlms et al. (2023) found that the gamification approach can also be used to convey information about the organization and the job during an assessment. To be more specific, in their study, among other game elements, Ohlms et al. embedded a traditional test in a job-related storyline in which avatars acted as potential colleagues and guided participants through the gamified ability test while providing applicants with information about the job and the organization. Through this specific gamification (i.e., simulation of a workday at the company), participants experienced significantly greater clarity about what to expect in the workplace compared to the non-gamified counterpart. Accordingly, a storified realistic test may offer the potential to improve participants' clarity about what to expect in a job by providing insights into the organizational culture, employees, or job tasks within the storified assessment. In this way, storification might also be used as a (realistic) job preview to reduce false expectations of applicants (Wanous, 1989). In contrast, a storified assessment that lacks job relevance and a traditional test that consists of abstract problem-solving tasks scarcely provide any reference point for participants about the job tasks and the organization. Thus, we hypothesize:

> **Hypothesis 5.** Clarity of work activity is higher for a storified realistic test than for a storified fantasy test and a non-storified test.

For applicants, the selection process is often their first encounter with an organization. Thus, in the early stages of a selection process, applicants usually have limited information about the organization, apart from the information they gather during the selection process (Turban, 2001). A theoretical framework that can be used to explain how applicants interpret information in situations where they have incomplete knowledge is signaling theory

(Spence, 1973). According to this theory, a signaling system comprises a sender, a receiver, and a message that is associated with unobservable attributes of the sender. When the sender and the receiver (i.e., recruiter and applicant) exchange information, the signaling system helps them to identify which information is valid for making decisions related to the job (Bangerter et al., 2012; Spence, 1973). In personnel selection research, signaling theory has been used to examine what conclusions applicants draw about unobservable symbolic attributes (e.g., modernity, innovativeness, competence, cf. Lievens & Highhouse, 2003) of an organization during the hiring process on the basis of observable attributes, such as selection instruments (e.g., Ehrhart & Ziegert, 2005; Wilhelmy et al., 2019). Specifically, in such situations, they might possibly interpret the information gathered as signals for organizational characteristics and working conditions. Consequently, using innovative storified assessments might signal to applicants that the organization itself is modern and cares about innovations (Georgiou & Lievens, 2022) whereas applicants might interpret the use of more traditional selection tools as signals of a more conservative culture (Lievens & Highhouse, 2003).

In contrast to fairness perceptions, the level of fantasy of a storified assessment may not impact the perceived modernity of the organization and enjoyment of the assessment. In fact, we posit that both a storified assessment requiring fighting aliens (i.e., storified fantasy assessment) and an assessment involving the management of an organization (i.e., storified realistic assessment) appear to be more modern and fun compared to a traditional computer-based ability test. In line with this, initial studies indicate that gamified assessments are perceived as more modern and fun than computer-based tests (Georgiou & Lievens, 2022; Harman & Brown, 2022; Ohlms et al., 2024b). Thus, we hypothesize:

> **Hypothesis 6.** (a) Modernity and (b) enjoyment are rated higher for both types of storified tests than for a non-storified test.

METHOD

Sample

We conducted a priori power analyses to determine the required sample size to test our hypotheses with a power of .80. The assumed effect sizes were based on previous storification research (e.g., Harman & Brown, 2022). Hypotheses 1, 2, and 5 required a sample size of $N = 159$ for a one-way analysis of variance (ANOVA) with three groups to detect a medium-sized effect of $f = .25$ (corresponding to a d of .50). Additionally, for the power analyses for the multivariate analyses of variance (MANOVAs) related to Hypotheses 3, 4 and 6, we assumed an effect size of $f^2 = .0625$ (also corresponding to a d of .50). Power analyses revealed a sample size of $N = 114$ for the overall multivariate test of Hypotheses 3 and 4 (three dependent variables) and of $N = 99$ for Hypothesis 6 (two dependent variables) but the follow-up one-way ANOVAs for the separate dependent variables again required an N of 159 with three groups to detect a medium-sized effect of $f = .25$. Accordingly, we aimed at collecting data from at least 159 participants.

We aimed to collect data from a heterogeneous participant pool that was more representative than purely student samples. Therefore, in Germany, participants were recruited using multiple recruitment methods. First, we gathered participants via different social media platforms (e.g., LinkedIn) and direct contact (e.g., direct e-mail inventions and personal networks), encouraging a diverse group of individuals to participate. Furthermore, psychology students

were able to register for the study and to receive course credit for their participation when they were members of the departmental participant pool at the first authors' university. In addition, participants were recruited through Survey Cycle (www.surveycircle.com), an online panel that allowed us to reach a broader audience beyond the university participant pool, social media, and direct contacts.

The initial sample consisted of 227 participants. However, prior to data analyses, 32 participants were excluded as they failed the attention check (six participants) and/or had an unrealistically short completion time, indicative of careless responding to the survey (28 participants). This led to a final sample of $N = 195$ participants. Of these, 53.8% identified as female, 45.6% as male, and 0.5% as diverse. Their age ranged from 17 to 66 with a mean of 34.62 years ($SD = 14.31$). For the highest educational degree, 5.1% reported an intermediate school-leaving certificate, 46.2% a general qualification for university entrance, 20.5% a bachelor's degree, 22.6% a master's degree, and 5.6% a PhD.

Procedure

The study was conducted as an online experiment using a between-subjects design. After giving their consent, all participants were presented with the same cover story: They were instructed to imagine that they had applied for a trainee program in process and quality management of renewable energies at the fictious organization CYQADELIC and had now been invited to take an online assessment. Subsequently, participants were randomly assigned to one of the three experimental groups—non-storified test, storified realistic test, or storified fantasy test (see below)—and were shown the respective tests.

Participants accessed the test corresponding to their experimental condition via a link in the survey that directed them to a website showing an excerpt from the respective online test. Thus, participants did not complete a full cognitive ability test, but took an excerpt from the assessment. This procedure aimed to allow participants to experience the actual selection test. Thus, participants completed five items from one subtest in which diagrams had to be analyzed, after which the instructions "In a real online test, 9 to 12 test items would follow now, with more time available for completion" were shown and "In a real test, there would be more than one test module. Another 3 to 6 test modules would follow here." After completing the excerpt of their respective test, participants answered items concerning different applicant reaction variables as well as demographic questions.[2]

Regarding the tests, the design of the website and the company logo was the same for all experimental conditions. Additionally, the excerpt from the actual cognitive ability test (i.e., the five items participants had to complete, which were diagram analysis tasks requiring participants to indicate whether a specific statement about the diagram was true, false, or indeterminate; see Figure 1(a)) was identical for all conditions. Thus, the three tests differed only regarding their storification. In particular, the non-storified test contained only the instructions (i.e., a welcome, as well as an explanation of test rules and control buttons) and the actual cognitive ability test (Figure 1(b)). In contrast, in the storified realistic test, the participants were guided through the assessment by potential colleagues (i.e., trainees; Figure 1(c)). These trainees told the participants that they already worked for the organization and asked the participants to support them with some tasks (i.e., the actual test). And in the storified fantasy test, participants were instructed that the company's headquarters had been taken over by aliens and were asked to help reconquer the office by completing the actual test (Figure 1(d)).

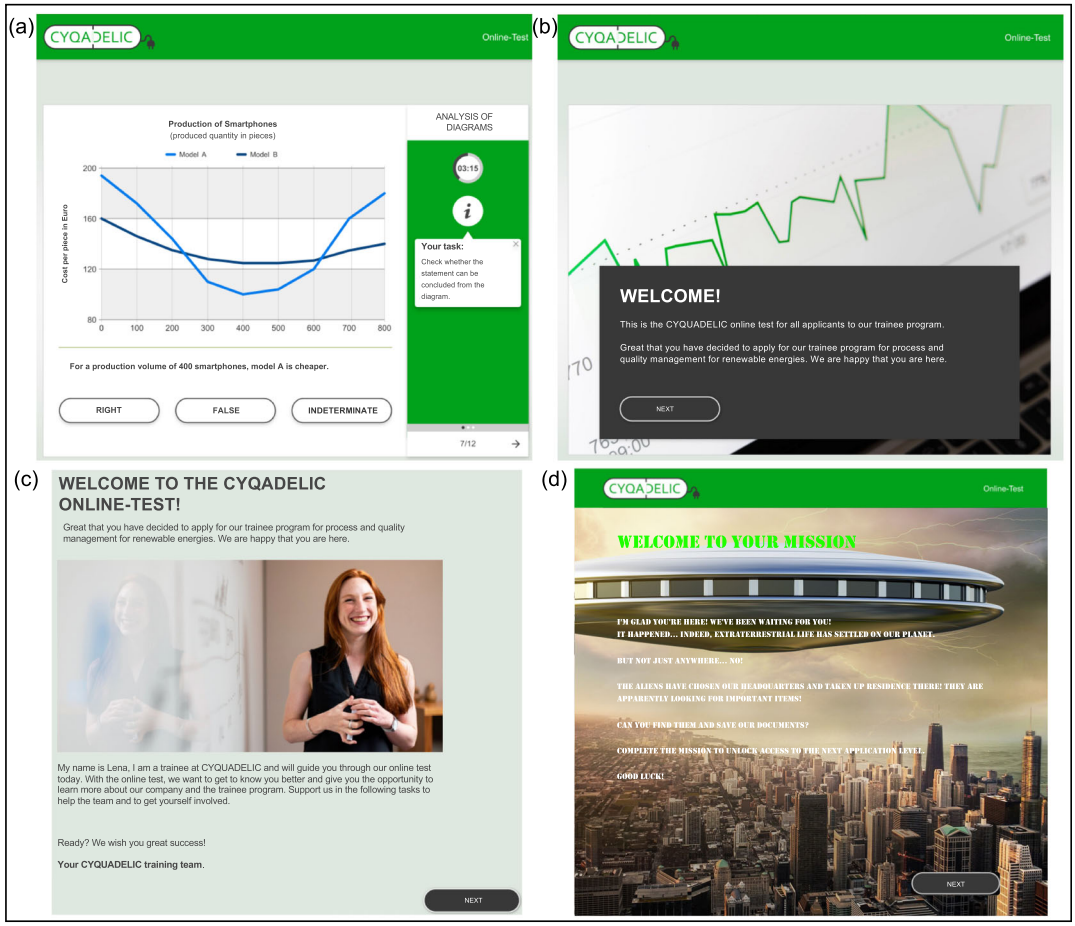

FIGURE 1 Examples of the different tests: example item of the actual cognitive ability test included in all tests (a), non-storified test (b), storified realistic test (c), and storified fantasy test (d). *Note*: The online assessment was presented to the participants in German. Copyright 2023 by CYQUEST GmbH. Adapted with permission.

To ensure that the non-storified test, which did not comprise a storyline, was as long as the storified assessments, the welcome, instruction, and farewell of the test were stretched over multiple pages. By doing so, we aimed to prevent the length of the test from influencing the results regarding applicants' reactions (Hausknecht et al., 2004; Ryan & Ployhart, 2000). Thus, we manipulated only the storyline, but kept the section of the cognitive ability test that participants completed constant across conditions. Accordingly, we aimed to hold all potential influencing factors (e.g., the actual test items, company logo, graphics) constant to ensure that any differences in applicant reactions could be attributed to the storyline itself (Lievens & Sackett, 2017).

Measures

Applicant reactions

The different applicant reaction variables were measured using established scales. Unless otherwise indicated, all items were rated on 5-point Likert scales ranging from 1 = *strongly disagree*

to 5 = *strongly agree*. The Appendix shows all items for the different applicant reaction variables and the respective item sources.

To assess perceived job-relatedness (two items; $\alpha = .91$), and procedural fairness (three items; $\alpha = .89$), we used two subscales of the Selection Procedural Justice Scale (Bauer et al., 2001). Organizational attractiveness ($\alpha = .91$) and behavioral intentions ($\alpha = .93$) were assessed using five items each from Highhouse et al. (2003). Clarity of work activity was measured using a three-item scale from Ohlms et al. (2023; $\alpha = .91$). Finally, the modernity of the organization was assessed using a semantic differential with two adjective pairs on a 5-point scale (conservative–innovative, traditional–modern; $\alpha = .88$) from Kanning et al. (2019) and enjoyment of the test ($\alpha = .86$) was measured with three items from Wilde et al. (2009).

We conducted a confirmatory factor analysis (CFA) with seven correlated factors to examine whether the different applicant reaction variables indeed represented separable constructs. This CFA fitted adequately, $\chi^2(209) = 407.41$, $p < .001$; CFI = .95; SRMR = .05; RMSEA = .07. In contrast, a single-factor model had a poor fit, $\chi^2(230) = 1344.82$, $p < .001$; CFI = .69, SRMR = .10, RMSEA = .16, and fitted significantly worse than the 7-factor model, $\Delta\chi^2(21) = 937.40$, $p < .001$. Moreover, we conducted a CFA with six correlated factors in which both fairness facets (i.e., perceived job-relatedness, procedural fairness) loaded on one factor, $\chi^2(215) = 584.52$, $p < .001$; CFI = .90, SRMR = .06, RMSEA = .10. This model fitted significantly worse than the 7-factor model, $\Delta\chi^2(6) = 177.10$, $p < .001$. However, owing to a rather high correlation between organizational attractiveness and behavioral intentions in the 7-factor model ($r = .91$, $p < .001$), we further tested a model with six correlated factors that defined organizational attractiveness and behavioral intentions as one factor, $\chi^2(215) = 412.53$, $p < .001$; CFI = .95, SRMR = .05, RMSEA = .07. This 6-factor model had adequate fit and did not differ significantly from the 7-factor model, $\Delta\chi^2(6) = 5.11$, $p = .53$. Consequently, we merged organizational attractiveness and behavioral intentions into the superordinate scale organizational attractiveness for all further analyses ($\alpha = .96$).[3]

Attention check

We included a code word at the end of each of the three test descriptions that participants had to enter before answering questions concerning their reactions to the test. This was to ensure that participants worked through to the end of the test excerpt and did not simply answer the applicant reactions items without going through the entire test excerpt. Participants who entered a wrong code word were removed from the data set.

RESULTS[4]

Table 1 presents means, standard deviations, intercorrelations, and reliabilities for the study variables. Across all groups, there was a negative correlation between age and gender ($r = -.16$, $p = .02$). Furthermore, before testing our hypotheses, we evaluated whether the three experimental groups differed regarding gender and age. A χ^2-test indicated no significant difference for gender, $\chi^2(2) = 0.87$, $p = .65$, and an ANOVA revealed no significant differences for age between the three groups, $F(2, 191) = 2.47$, $p = .09$.

TABLE 1 Descriptive information and correlations for the study variables.

Variable	*M*	*SD*	1	2	3	4	5	6	7	8
1. Age[a]	34.62	14.31	(−)							
2. Gender[a]	0.54	0.50	−.16*	(−)						
3. Perceived job-relatedness	2.66	1.22	−.18*	.08	(.91)					
4. Procedural fairness	3.09	1.05	−.27***	.14*	.51***	(.89)				
5. Organizational attractiveness	3.24	0.97	−.17*	.13	.55***	.63***	(.91)			
6. Clarity of work activity	2.02	1.02	−.14	.14*	.56***	.52***	.58***	(.91)		
7. Modernity	3.91	0.91	−.17*	.15*	.15*	.14*	.37***	.32***	(.88)	
8. Enjoyment	3.59	0.95	−.14*	.18*	.26***	.36***	.50***	.29***	.33***	(.86)

Note: $N = 195$. Reliabilities (internal consistencies) are shown in the diagonal. Gender is coded 0 = male, 1 = female. [a]$n = 194$. *$p < .05$, and ***$p < .001$. All correlations were tested two-tailed.
Abbreviation: SD, standard deviation.

Testing of hypotheses

Hypothesis 1 posited that perceived job-relatedness is higher for the storified realistic test than for the storified fantasy test, and Hypothesis 2 predicted that perceived job-relatedness is lower for the non-storified test than for the storified realistic test, but higher than for the storified fantasy test. The means followed the predicted pattern, and the one-way ANOVA conducted to test these hypotheses was also significant, $F(2, 192) = 11.90$, $p < .001$, $\eta^2 = .11$ (see Table 2 and Figure 2). Furthermore, results for post-hoc tests using the Scheffé-procedure confirmed that the storified realistic test was rated significantly higher on perceived job-relatedness than the storified fantasy test ($d = 0.80$, $p < .001$) as well as the non-storified test ($d = 0.56$, $p = .007$). However, there was no significant difference between the non-storified and storified fantasy test ($d = 0.25$, $p = .38$). Thus, Hypothesis 1 was supported, whereas Hypothesis 2 was partly supported by these results.

The descriptive pattern was also in line with Hypothesis 3, predicting that (a) procedural fairness and (b) organizational attractiveness are higher for the storified realistic test than for the storified fantasy test as well as with Hypothesis 4, postulating that (a) procedural fairness and (b) organizational attractiveness are lower for the non-storified test than for the storified realistic test, but higher than for the storified fantasy test. To test these hypotheses, we conducted a one-way MANOVA. Results revealed a significant multivariate effect, Wilks' $\lambda = 0.82$, $F(4, 382) = 10.27$, $p < .001$, $\eta^2 = .10$. As shown in Table 2, separate one-way ANOVAs also revealed significant effects for the two applicant reaction variables, both Fs > 11.73, both ps < .001. Subsequent post-hoc tests using the Scheffé-procedure showed support for Hypotheses 3. More precisely, the storified realistic test was perceived more positively regarding procedural fairness ($d = 0.81$, $p < .001$), and organizational attractiveness ($d = 0.80$, $p < .001$). Furthermore, Hypothesis 4 was partly supported by these analyses. Specifically, there was a large positive difference for organizational attractiveness ($d = 0.92$, $p < .001$) between the storified fantasy test vs. the non-storified test, but no significant differences regarding these two assessments for procedural fairness ($d = 0.37$, $p = .14$). Additionally, procedural fairness ($d = 0.46$, $p = .03$) was higher for the non-storified test than for the storified fantasy test, but there was no significant difference regarding organizational attractiveness ($d < 0.01$, $p = 1.00$).

TABLE 2 Means and *SD*s for the experimental groups and results of the one-way ANOVAs for the different applicant reaction variables as dependent variables and the selection instrument as independent variable.

	Selection instrument							
	Storified realistic ($n = 72$)	Non-storified ($n = 59$)	Storified fantasy ($n = 64$)	ANOVA		Cohen's *d*s and results for post-hoc tests		
Dependent variable	*M (SD)*	*M (SD)*	*M (SD)*	$F(2, 192)$	η^2	Storified realistic vs. non-storified[1]	Storified realistic vs. storified fantasy[1]	Non-storified vs. storified fantasy[2]
Perceived job-relatedness	3.17_a (1.18)	2.52_b (1.11)	2.23_b (1.16)	11.90***	.11	0.56**	0.80***	0.25
Procedural fairness	3.47_a (0.93)	3.12_a (0.97)	2.64_b (1.11)	11.73***	.11	0.37	0.81***	0.46*
Organizational attractiveness	3.71_a (0.77)	2.97_b (0.85)	2.96_b (1.09)	15.05***	.14	0.92***	0.80***	0.01
Clarity of work activity	2.59_a (1.05)	1.72_b (0.74)	1.64_a (0.92)	22.58***	.19	0.94***	0.96***	0.10
Modernity	4.21_a (0.63)	3.10_b (0.96)	4.30_a (0.63)	49.71***	.34	1.40***	−0.14	−1.49***
Enjoyment	$3.64_{a,b}$ (0.93)	3.33_a (0.98)	3.79_b (0.90)	3.80*	.04	0.33	−0.16	−0.49*

Note: Means sharing a common subscript did not differ significantly according to Schéffe post-hoc tests. [1]Positive values for Cohen's *d* indicate higher means in the storified realistic condition. [2]Positive values for Cohen's *d* indicate higher means in the non-storified condition.

*$p < .05$, **$p < .01$, and ***$p < .001$.

Abbreviations: ANOVA, analysis of variance; SD, standard deviation.

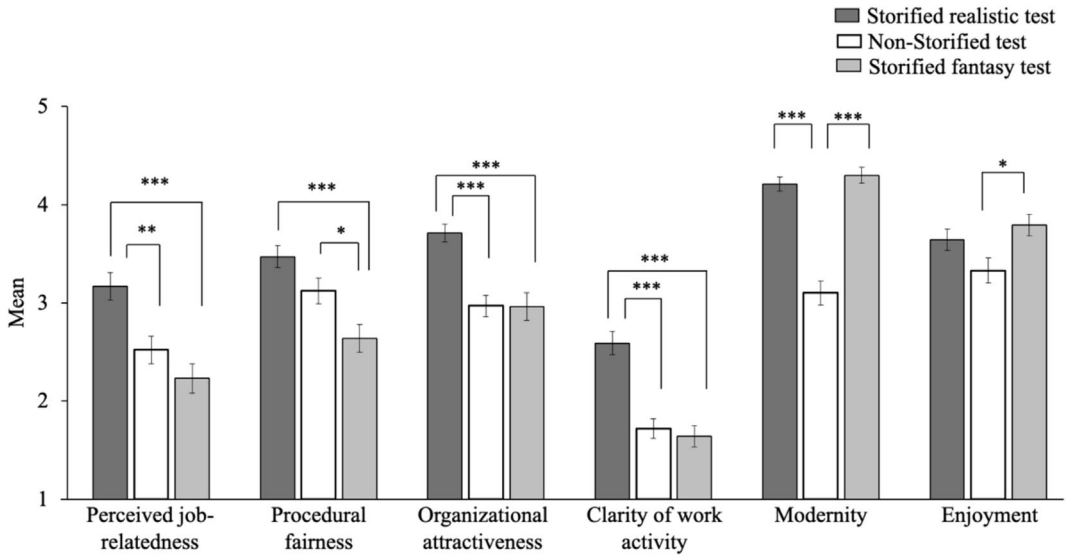

FIGURE 2 Results for the different applicant reaction variables as dependent variables and the selection instrument as the independent variable. *Note:* Error bars represent standard errors of the mean. Significance levels refer to results from Schéffe post-hoc tests. $n_{\text{storified realistic}} = 72$, $n_{\text{non-storified}} = 59$, and $n_{\text{storified fantasy}} = 64$. $^{*}p < .05$, $^{**}p < .01$, and $^{***}p < .001$.

A potential shortcoming of the different (M)ANOVAs and the corresponding post-hoc tests that we conducted to evaluate Hypotheses 1 to 4 is that they represent nondirectional tests. Therefore, we decided also to test our hypotheses regarding the fairness variables from Gilliland's (1993) fairness model (i.e., perceived job-relatedness, procedural fairness) with more powerful contrast tests (Rosenthal & Rosnow, 1985). Accordingly, we used contrast weights of +1, 0, and −1 for the storified realist test, the non-storified test, and storified fantasy test to represent the predicted (and preregistered order) of the means. In line with the descriptive pattern across the three experimental groups (see Table 2), the results from the contrast tests for perceived job-relatedness and procedural fairness, unequivocally confirmed the predicted pattern, both ts > 11.73, both ps < .001. Thus, with this more powerful statistical approach, Hypotheses 1 to 4 were all supported.

Hypothesis 5 assumed that clarity regarding work activity is higher for the storified realistic test than for the other two assessments. To test this assumption, we used a one-way ANOVA and found a significant effect, $F(2, 192) = 22.58$, $p < .001$, $\eta^2 = .11$. In line with Hypothesis 5, the post-hoc test using the Scheffé-procedure found large positive effects for the perceived clarity of work activity of the storified realistic test in comparison the storified fantasy test ($d = 0.96$) as well as the non-storified assessment ($d = 0.94$, both ps < .001; see Table 2).

Finally, to examine Hypothesis 6, postulating that (a) modernity and (b) enjoyment are rated higher for both of the storified tests than for the non-storified test, we conducted a MANOVA. Results showed a significant multivariate effect, Wilks' $\lambda = 0.66$, $F(4, 382) = 22.33$, $p < .001$, $\eta^2 = .19$, as well as significant effects for the separate one-way ANOVAs, both Fs > 3.80, both ps < .02. In line with Hypothesis 6a, post-hoc tests using the Scheffé-procedure revealed that both storified assessments were rated higher on modernity than the non-storified test (both ds > 1.40, $p < .001$; see Table 2). Furthermore, there was partial support for Hypothesis 6b. Specifically, consistent with our assumption, enjoyment was rated higher for the storified

fantasy test than for the non-storified test ($d = 0.49$, $p = .03$), but there was no significant difference between the storified realistic test and the non-storified test ($d = 0.33$, $p = .17$).

DISCUSSION

Although technology-based selection instruments such as storified assessments involve many design choices, there is a surprising lack of agreement and recommendations regarding the impact of these design features. The present study focused on one specific design consideration, namely, a storyline and its level of fantasy as well as its effect on applicant reactions. There are opposing views as to whether the storyline should reflect actual work activities or not. One perspective argues that fantasy storylines, such as the task of fighting aliens, more closely match modern commercially available games and should therefore be more enjoyable for applicants, thus possibly creating a sense of immersion in the game. As a flip side, storylines that are totally unrelated to the job in question might lead applicants to wonder why they must take such a test to get a job offer, which could negatively impact their reactions. Therefore, we suggested that a realistic storyline in a game, in addition to having a playful element, also provides a preview of the job and of the organization during the assessment.

The present study compared these two perspectives by developing and examining hypotheses that build on Gilliland's (1993) fairness model and on signaling theory (Spence, 1973), suggesting that the type of fantasy in a storyline within a storified assessment influences applicant reactions. To test our hypotheses, we compared reactions to a realistic storified, a fantasy storified, and a non-storified test. There was general support for our hypotheses. Thus, the level of fantasy in a storyline within a storified assessment plays an important role for applicant reaction variables, whereby the storified realistic test was rated most favorably, followed by the non-storified test, and finally the storified fantasy assessment.

Specifically, the storified realistic test was perceived more positively than the storified fantasy assessment on job-relatedness, procedural fairness, organizational attractiveness, and perceived clarity of job activity. Secondly, the storified fantasy test led to applicants' perceiving the assessment as more enjoyable and the organization as more modern compared to its non-storified counterpart, but the storified fantasy test did not positively affect organizational attractiveness. In fact, the storified fantasy test negatively influenced the perceived procedural fairness compared to a non-storified test. In contrast, we found a large positive effect of the storified realistic assessment on organizational attractiveness, clarity of work activity, and perceived modernity of the organization as well as a medium-sized effect on perceived job-relatedness relative to the non-storified test. Below, we discuss the implications of our findings for research and practice.

Implications for gamification research and theory

One implication for gamification research lies in the fact that previous studies on gamified assessments have yielded mixed results to date, suggesting that the gamification of a traditional test may not automatically result in improved applicant reactions (see Ramos-Villagrasa et al., 2022, for a review). A potential explanation for the divergent findings concerning applicant reactions is the variety of manifestations that gamified assessments can take on depending on their specific design. However, given that the gamified assessments in previous studies have

tended to conflate various game elements (Georgiou & Nikolaou, 2020; Landers et al., 2022), we do not know which of these elements have a particularly positive or negative effect. Hence, there has been a recent call for research on determinants that might substantially positively or negatively impact applicant reactions to gamified procedures (Ramos-Villagrasa et al., 2022).

In this context, our results extend our knowledge of the use of storylines, more precisely, the level of fantasy of a storyline as one specific gamification element. Specifically, we isolated the effect of the type of fantasy of a storyline as a method factor (Lievens & Sackett, 2017), holding other potential influencing factors constant (e.g., the actual assessment, company logo, graphics). By doing so, our results provide important insights for research on this gamification element (i.e., the level of fantasy of a storyline). Accordingly, we found that the level of fantasy is an important aspect in the design of a gamified assessment, through which applicant reactions might be affected considerably. Thus, our results align with the findings of previous research, which has shown that selection instruments that simulate everyday working life, such as work samples or situational interviews, or that embed test items in a business context (i.e., that all have a clear job-relatedness) are generally associated with more positive applicant reactions than selection instruments with lower job-relatedness (e.g., noncontextualized personality tests or cognitive ability tests; Bauer et al., 2001; Hausknecht et al., 2004; Rynes & Connerley, 1993). However, in contrast to a work sample, the actual test items used in our test did not reflect real work tasks. Thus, only the storyline of the storified realistic test itself was intended to reflect an actual work situation. As a result, the storified realistic test used in our study may differ from a work sample since the test items do not contain content that is sampled from work. Instead, the storified realistic test was just dressed up like a work sample, whereas a real work sample contains actual work tasks. Furthermore, it may be interesting to examine whether applicants' long-term reactions differ when they later discover that the actual job does not involve solving such problems as those presented in the actual test items in our storified realistic test.

Another significant finding was that a job-related or realistic storification, in comparison to a storified fantasy and non-storified test, also provides the opportunity to convey some kind of job preview of the actual work activity and the organization during the assessment (Wanous, 1989). This is important given that previous research on gamified assessments has so far primarily focused on the variables from Gilliland's (1993) fairness model as well as on attitudes towards the organization (e.g., Ellison et al., 2020; Georgiou & Nikolaou, 2020; Ohlms et al., 2024a). However, little attention has been paid to whether the gamification or storification approach might also provide other benefits beyond the enhancement of fairness perceptions. Accordingly, in line with the results of Ohlms et al. (2023), our findings show that the storification approach, specifically the use of a realistic storyline, may also be used to present a job preview during the assessment.

Implications for practice

For practice, our results also suggest several recommendations. First, the present study indicates that storification and/or gamification per se does not automatically lead to improved applicant reactions in comparison with a traditional computer-based test. Instead, our results highlight that the storyline's level of fantasy can substantially influence various applicant reactions and thus should be considered (in addition to other outcome variables such as validity or reliability) when developing a storified assessment or when buying such an assessment from a provider.

Secondly, our findings offer advice to practitioners and researchers who are facing difficulties in deciding which type of storyline to choose—a fantasy or a realistic storyline—when designing or buying a storified/gamified assessment. Although both a fantasy and a realistic storification may improve the perceived modernity of the organization relative to a non-storified assessment, a fantasy storyline will probably not lead to applicants perceiving the organization to be more attractive and being more willing to accept a job offer. In fact, a fantasy storyline may even reduce the perceived fairness of the assessment compared to a traditional test. In contrast, a realistic storyline offers the opportunity both to improve the perceived job-relatedness of a test and to enhance applicants' attitudes towards the organization.

Consequently, our findings suggest that organizations aiming to ensure that their assessment is perceived as fun and their organization as modern may rely on both fantasy and realistic storylines within a storified assessment. However, when simultaneously seeking to improve the perceived fairness of an assessment, as well as organizational attractiveness and applicants' intentions to accept a job offer, compared to a traditional test, we recommend the choice of a realistic storyline when both assessments have comparable psychometric properties (i.e., validity and reliability).

Limitations and directions for future research

One limitation concerns the fact that we compared only two potential levels of fantasy of a storified assessment, that means, high fantasy and no fantasy, with a non-storified assessment. However, when designing a storified assessment, a wide variety of different degrees of fantasy in the storyline could be used. Accordingly, it would be interesting to further examine what degree of fantasy leads to the most positive applicant reactions.

Secondly, and related to the previous limitation, we focused on the effect of storification on applicant reactions to cognitive ability tests, manipulating only the level of fantasy of the storified assessment but not the degree of storification itself. Thereby, we chose a mild degree of storification by only embedding a traditional non-gamified cognitive ability test into a storyline without storifying the actual test items. In general, gamified assessments can vary considerably in their degree of gamification. At one end of the continuum, "an existing cognitive ability test might be gamified by simply presenting it as if it were a game, a shallow type of gamification called framification or game-framing" (Landers & Sanchez, 2022, p. 2). At the other end of the continuum, there are gamified assessments such as Owiwi (see Georgiou et al., 2019), which use various game elements such as a storyline, avatars, and a visual progression bar. Thus, we chose one of many possible degrees of storification with our low to medium degree of storification. Nonetheless, this exerted significant effects on applicant reactions with moderate to large effects sizes for most comparisons. Future research should examine how different degrees of storification (e.g., storifying the actual test items vs. only embedding the test items in a storyline) affect applicant reactions and performance.

As a third limitation, we focused on the effect of the game element storification on applicant reactions; however, there are many other game elements, such as points, leaderboards, or levels that may also influence applicant reactions. Therefore, more research is needed on the effects of other specific game elements. Future research should also examine the effects of specific game elements (e.g., for levels), in isolation in the context of personnel selection to assess how such game elements affect applicant reactions.

Furthermore, we examined only how the level of fantasy of a storified test affects applicant reactions. However, gains in applicant reactions are only worthwhile if the psychometric properties (i.e., reliability and validity) of the test do not suffer from the changes at the same time. Thus, further research should investigate whether the level of fantasy of a storyline also influences the validity of the assessment, or whether different applicant groups are targeted more or less effectively depending on the level of fantasy. In general, more research is needed concerning the validity, usefulness, and potential subgroup differences in performance and acceptance of specific game elements (Ramos-Villagrasa et al., 2022).

A final limitation of our study is that the participants did not take the entire test. We do not know how the effects of the storyline develop over a longer period of time. In other contexts (e.g., education), some studies suggest that the positive effects of gamification may be due only to the novelty effect, which may decrease after a certain period of time (e.g., Hamari et al., 2014; Rodrigues et al., 2022). However, participants had to complete an excerpt of the test and thus had the chance to experience it. Furthermore, our study was not conducted in a high-stakes situation. Instead, people had to imagine applying for an attractive job, which is a common approach in applicant reaction research (e.g., Georgiou & Nikolaou, 2020; Landers et al., 2022). Accordingly, it would be interesting for future research to re-examine the effects of fantasy vs. realistic storylines by having actual applicants complete the entire storified test and to see whether the effects of the storyline become weaker when applicants complete more than a few items.

CONCLUSION

This study has extended our knowledge of the use of storylines as one specific gamification element by isolating the effect of the level of fantasy of a storified assessment on reactions towards it. We found that the level of fantasy of a storified test considerably affects test taker reactions. Contrary to propositions from some providers of gamified tests as well as to suggestions in the literature (e.g., see Bhatia & Ryan, 2018; Fetzer et al., 2017), storification of a traditional, non-storified test does not automatically lead to improved perceptions of fairness, organizational attractiveness, and intentions to accept a job offer among potential applicants. Concerning the question of whether a fantasy or a realistic storyline within a storified assessment should be chosen in preference to a traditional assessment in order to positively influence applicant reactions, we found that a realistic storification is to be preferred when the assessment has comparable psychometric properties (i.e., validity and reliability). Furthermore, in addition to improving fairness perceptions and attitudes towards the organization compared to a storified fantasy test, another advantage of a realistic storyline is the opportunity to convey a job preview of the actual work activity as well as the organization during the assessment.

ACKNOWLEDGMENTS

The authors would like to thank Lara Schmid for her help concerning data collection and Joachim Diercks and his team at CYQUEST for their help in generating and programming the study materials. A slightly modified version of this article is part of the first author's dissertation: Ohlms, M. L. (2024). Digitization in Personnel selection: Adaptation and enhancement of traditional personnel selection instruments using gamification and game-based assessment [Doctoral dissertation, Universität Ulm]. kiz Universität Ulm. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in OSF at https://osf.io/x2mgy/.

ETHICS STATEMENT

Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

ORCID

Marie L. Ohlms https://orcid.org/0000-0003-4022-6904
Klaus G. Melchers https://orcid.org/0000-0003-4211-6450
Filip Lievens https://orcid.org/0000-0002-9487-5187

ENDNOTES

[1] This study was pre-registered at: https://aspredicted.org/TJG_6QP. However, note that in comparison to the pre-registration, some of the labels changed during the revision process.

[2] In addition, we gathered information on participants' video game experience, computer self-efficacy, and openness to experience. However, we decided not to report these variables below as they were not the focus of interest for the present study.

[3] We also conducted the subsequent analyses related to organizational attractiveness and behavioral intentions without combining these two scales into one. The results for analyzing organizational attractiveness and behavioral intentions separately vs. combined into one scale did not change the results meaningfully and led to identical conclusions.

[4] Results in this section can be reproduced using the data set and code that can be found at https://osf.io/x2mgy/.

REFERENCES

Akkerman, S., Admiraal, W., & Huizenga, J. (2009). Storification in history education: A mobile game in and about medieval Amsterdam. *Computers & Education*, *52*(2), 449–459. https://doi.org/10.1016/j.compedu.2008.09.014

Aldemir, T., Celik, B., & Kaplan, G. (2018). A qualitative investigation of student perceptions of game elements in a gamified course. *Computers in Human Behavior*, *78*, 235–254. https://doi.org/10.1016/j.chb.2017.10.001

Bai, S., Hew, K. F., Gonda, D. E., Huang, B., & Liang, X. (2022). Incorporating fantasy into gamification promotes student learning and quality of online interaction. *International Journal of Educational Technology in Higher Education*, *19*, 29. https://doi.org/10.1186/s41239-022-00335-9

Bangerter, A., Roulin, N., & König, C. J. (2012). Personnel selection as a signaling game. *Journal of Applied Psychology*, *97*(4), 719–738. https://doi.org/10.1037/a0026078

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

Bedwell, W. L., Pavlas, D., Heyne, K., Lazzara, E. H., & Salas, E. (2012). Toward a taxonomy linking game attributes to learning: An empirical study. *Simulation & Gaming*, *43*(6), 729–760. https://doi.org/10.1177/1046878112439444

Bhatia, S., & Ryan, A. M. (2018). Hiring for the win: Game-based assessment in employee selection. In J. H. Dulebohn & D. L. Stone (Eds.), *The brave new world of eHRM 2.0* (pp. 81–110). Information Age Publishing.

Chapman, D. S., Uggerslev, K. L., Carroll, S. A., Piasentin, K. A., & Jones, D. A. (2005). Applicant attraction to organizations and job choice: A meta-analytic review of the correlates of recruiting outcomes. *Journal of Applied Psychology*, *90*(5), 928–944. https://doi.org/10.1037/0021-9010.90.5.928

Ehrhart, K. H., & Ziegert, J. C. (2005). Why are individuals attracted to organizations? *Journal of Management*, *31*(6), 901–919. https://doi.org/10.1177/0149206305279759

Ellison, L. J., McClure Johnson, T., Tomczak, D., Siemsen, A., & Gonzalez, M. F. (2020). Game on! Exploring reactions to game-based selection assessments. *Journal of Managerial Psychology*, *35*(4), 241–254. https://doi.org/10.1108/JMP-09-2018-0414

Fetzer, M., McNamara, J., & Geimer, J. L. (2017). Gamification, serious games and personnel selection. In H. W. Goldstein, E. D. Pulakos, J. Passmore, & C. Semedo (Eds.), Wiley Blackwell handbook of the psychology of recruitment, selection and employee retention (pp. 293–309). Wiley. https://doi.org/10.1002/9781118972472

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

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work?—A literature review of empirical studies on gamification. In *47th Hawaii international conference on system sciences* (pp. 3025-3034). https://doi.org/10.1109/HICSS.2014.377

Harman, J. L., & Brown, K. D. (2022). Illustrating a narrative: A test of game elements in game-like personality assessment. *International Journal of Selection and Assessment*, *30*(1), 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

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

Kanning, U. P., Kraul, L.-F., & Litz, R. Z. (2019). Einstellungen zu digitalen Methoden der Personalauswahl [Attitudes towards digital personnel selection methods]. *Journal of Business and Media Psychology*, *10*(1), 57–61. https://journal-bmp.de/wp-content/uploads/8a_Einstellungen-zu-digitalen-Methoden-der-Personalauswahl05.02.2020-1.pdf

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/apl0000954

Landers, R. N., & Collmus, A. B. (2022). Gamifying a personality measure by converting it into a story: Convergence, incremental prediction, faking, and reactions. *International Journal of Selection and Assessment*, *30*(1), 145–156. https://doi.org/10.1111/ijsa.12373

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

Lievens, F., & Sackett, P. R. (2017). The effects of predictor method factors on selection outcomes: A modular approach to personnel selection procedures. *Journal of Applied Psychology*, *102*(1), 43–66. https://doi.org/10.1037/apl0000160

Malone, T. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science*, *5*(4), 333–369. https://doi.org/10.1016/S0364-0213(81)80017-1

Malone, T. W., & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning and instruction* (pp. 223–253). Erlbaum.

McCord, J. L., Harman, J. L., & Purl, J. (2019). Game-like personality testing: An emerging mode of personality assessment. *Personality and Individual Differences*, *143*(1), 95–102. https://doi.org/10.1016/j.paid.2019.02.017

Melchers, K. G., & Basch, J. M. (2022). Fair play? Sex-, age-, and job-related correlates of performance in a computer-based simulation game. *International Journal of Selection and Assessment*, *30*(1), 48–61. https://doi.org/10.1111/ijsa.12337

Ohlms, M. L., Melchers, K. G., & Kanning, U. P. (2024a). Can we playfully measure cognitive ability? Construct-related validity and applicant reactions. *International Journal of Selection and Assessment*, *32*(1), 91–107. https://doi.org/10.1111/ijsa.12450

Ohlms, M. L., Melchers, K. G., & Kanning, U. P. (2024b). Playful personnel selection: The use of traditional vs. game-related personnel selection methods and their perception from the recruiters' and applicants' perspectives. *International Journal of Selection and Assessment*, *32*(3), 381–398. https://doi.org/10.1111/ijsa.12466

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

Parker, L. E., & Lepper, M. R. (1992). Effects of fantasy contexts on children's learning and motivation: Making learning more fun. *Journal of Personality and Social Psychology*, *62*(4), 625–633. https://doi.org/10.1037/0022-3514.62.4.625

Ramos-Villagrasa, P. J., Fernández-Del-Río, E., & Castro, Á. (2022). Game-related assessments for personnel selection: A systematic review. *Frontiers in Psychology*, *13*, 952002. https://doi.org/10.3389/fpsyg.2022.952002

Reddock, C. M., Auer, E. M., & Landers, R. N. (2020). A theory of branched situational judgment tests and their applicant reactions. *Journal of Managerial Psychology*, *35*(4), 255–270. https://doi.org/10.1108/JMP-10-2018-0434

Rodrigues, L., Pereira, F. D., Toda, A. M., Palomino, P. T., Pessoa, M., Carvalho, L. S. G., Fernandes, D., Oliveira, E. H. T., Cristea, A. I., & Isotani, S. (2022). Gamification suffers from the novelty effect but benefits from the familiarization effect: Findings from a longitudinal study. *International Journal of Educational Technology in Higher Education*, *19*, 13. https://doi.org/10.1186/s41239-021-00314-6

Rosenthal, R., & Rosnow, R. L. (1985). *Contrast analysis: Focused comparisons in the analysis of variance*. Cambridge University Press.

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

Rynes, S. L., & Connerley, M. L. (1993). Applicant reactions to alternative selection procedures. *Journal of Business and Psychology*, *7*(3), 261–277. https://doi.org/10.1007/BF01015754

Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review*, *32*(1), 77–112. https://doi.org/10.1007/s10648-019-09498-w

Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics*, *87*(3), 355–374. https://doi.org/10.2307/1882010

Toda, A. M., Klock, A. C., Oliveira, W., Palomino, P. T., Rodrigues, L., Shi, L., Bittencourt, I., Gasparini, I., Isotani, S., & Cristea, A. I. (2019). Analysing gamification elements in educational environments using an existing gamification taxonomy. *Smart Learning Environments*, *6(1)*, 16. https://doi.org/10.1186/s40561-019-0106-1

Turban, D. B. (2001). Organizational attractiveness as an employer on college campuses: An examination of the applicant population. *Journal of Vocational Behavior*, *58*(2), 293–312. https://doi.org/10.1006/jvbe.2000.1765

Wanous, J. P. (1989). Installing a realistic job preview: Ten tough choices. *Personnel Psychology*, *42*(1), 117–134. https://doi.org/10.1111/j.1744-6570.1989.tb01553.x

Wilde, M., Bätz, K., Kovaleva, A., & Urhahne, D. (2009). Überprüfung einer Kurzskala intrinsischer Motivation [Validation of a short scale of intrinsic motivation]. *Zeitschrift für Didaktik der Naturwissenschaften*, *15*, 31–45. https://pub.uni-bielefeld.de/record/2404161

Wilhelmy, A., Kleinmann, M., Melchers, K. G., & Lievens, F. (2019). What do consistency and personableness in the interview signal to applicants? Investigating indirect effects on organizational attractiveness through

symbolic organizational attributes. *Journal of Business and Psychology*, *34*(5), 671–684. https://doi.org/10.1007/s10869-018-9600-7

Woods, S. A., Ahmed, S., Nikolaou, I., Costa, A. C., & Anderson, N. R. (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

How to cite this article: Ohlms, M. L., Melchers, K. G., & Lievens, F. (2024). It's just a game! Effects of fantasy in a storified test on applicant reactions. *Applied Psychology*, 1–21. https://doi.org/10.1111/apps.12569

APPENDIX

Items used to measure applicant reactions to the different tests.

Scale	Items used in the current study	Source
Perceived job-relatedness[b]	It would be clear to anyone that this test is related to the job as a trainee. The content of the test was clearly related to the job as a trainee.	Bauer et al. (2001)
General procedural fairness[b]	I think that this test is a fair way to select people for the job as a trainee. I think that the test itself is fair. Overall, the method used was fair.	Bauer et al. (2001)
Organizational attractiveness[b]	For me, this company would be a good place to work. I would not be interested in this company except as a last resort.[a] This company is attractive to me as a place for employment. I am interested in learning more about this company. A job at this company is very appealing to me.	Highhouse et al. (2003)
Behavioral Intentions[b]	I would accept a job offer from this company. I would make this company one of my first choices as an employer. If this company invited me for another job interview after this procedure, I would go. I would exert a great deal of effort to work for this company. I would recommend this company to a friend looking for a job.	Highhouse et al. (2003)
Clarity of work activity[b]	I have a clear idea of what it is like to work at the company CYQADELIC. I know which work tasks would be expected of me in the job as a trainee. I have a clear idea of what the daily work routine at the company CYQADELIC would be like.	Ohlms et al. (2023)
Modernity[c]	Conservative–innovative Traditional–modern	Kanning et al. (2019)
Enjoyment[b]	I enjoyed this test. I find this test to be very interesting. I find this test to be entertaining	Wilde et al. (2009)

Note: [a]Reverse coded. [b]Items were presented using a 5-point rating scale (1 = *strongly disagree* to 5 = *strongly agree*). [c]Items were captured using a semantic differential with two adjective pairs on a 5-point rating scale. All items were presented in German.