Received: 5 April 2018 | Revised: 22 March 2019 | Accepted: 25 March 2019

DOI: 10.1111/ijsa.12240

FEATURE ARTICLE

International Journal of Selection and Assessment | WILEY

Gamification in employee selection: The development of a gamified assessment

Konstantina Georgiou | Athanasios Gouras | Ioannis Nikolaou

Department of Management Science and Technology, School of Business, Athens University of Economics and Business, Athens, Greece

Correspondence
Konstantina Georgiou, Department of Management Science and Technology, School of Business, Athens University of Economics and Business, 76, Patission Str., Athens GR10434, Greece.
Email: kongeorgiou@aueb.gr

Abstract

Gamification has attracted increased attention among organizations and human resource professionals recently, as a novel and promising concept for attracting and selecting prospective employees. In the current study, we explore the construct validity of a new gamified assessment method in employee selection that we developed following the situational judgement test (SJT) methodology. Our findings support the applicability of game elements into a traditional form of assessment built to assess candidates' soft skills. Specifically, our study contributes to research on gamification and employee selection exploring the construct validity of a gamified assessment method indicating that the psychometric properties of SJTs and their transformation into a gamified assessment are a suitable avenue for future research and practice in this field.

KEYWORDS
employee selection, gamified assessment method, situational judgement test

1 | INTRODUCTION

New technologies, such as gamification, game-based assessments, and serious games, have recently attracted increased attention in the field of talent identification (Chamorro-Premuzic, Akhtar, Winsborough, & Sherman, 2017). Serious games are games designed and used for a primary goal other than entertainment (Michael & Chen, 2005). In turn, gamification refers to the incorporation of game elements into nongaming activities in any context, such as the workplace, giving birth to game-based assessments, which can be classified according to the level of game characteristics they employ from gamified assessments, such as multimedia situational judgement test (SJTs) to different styles of games, such as Candy Crush and Flight Simulator (Hawkes, Cek, & Handler, 2017).

Gamification has been applied to employee selection settings in order to make assessment methods more game-like improving thus applicant reactions and possibly increase the prediction of job performance (Armstrong, Ferrell, Collmus, & Landers, 2016a). However, no published studies to date have established the effectiveness of gamification in the recruitment and selection process. Therefore, a question arises. Should researchers and professionals in Work/Organizational Psychology and Human Resource Management be interested in the use and effectiveness of gamified selection methods? Gamified selection methods might improve hiring decisions. For example, traditional methods used in employee selection make two inferential leaps; one about applicants' rating on multiple-choice items measuring traits and competencies and the extent to which they possess these traits or competencies, and another between competencies and applicants' actual job performance, which gamified selection methods may not make (Fetzer, McNamara, & Geimer, 2017). Playing online gamified assessments might simulate situations where individuals' intentions and behaviors are shown. Depending on the type of the game design and elements used in assessments, applicants' attention might be drifted from the fact that they are evaluated, showcasing thus their true behaviors and, as a result, reduce faking and/or social desirability biases (Armstrong, Landers, & Collmus, 2016b). Therefore, gamified selection methods might reduce the traditional methods' inferential leaps, improving thus the prediction of job performance.

Since recent studies have dictated the applicability of SJTs in high fidelity modes, such as video, multimedia, and interactive formats (Lievens & Sackett, 2006a) and gamified contexts (Armstrong, Landers et al., 2016b), the purpose of our research is to explore the development and construct validity of a SJT assessment that has been subsequently converted into a gamified assessment. Specifically, we used gamification to gamify a previous form of assessment that we initially developed (SJT). To achieve our goal, we conducted two studies; the development and construct validity of a SJT (Study 1), and the replication of the results to a gamified version of the SJT and its cross-validation (Study 2).

2 | GAMIFICATION IN EMPLOYEE SELECTION

Similarly to work sample and multimedia assessment tools, gamified selection methods assess applicants' knowledge, skills, abilities, and other characteristics (KSAOs), which are supported to predict job performance (e.g., Lievens & De Soete, 2012; Schmidt & Hunter, 1998). Moreover, the use of gamified selection methods might lead to increased engagement and positive perceptions of the organization signaling that it is at the cutting edge of technology offering competitive advantage in the war of talent (Fetzer et al., 2017). Chow and Chapman (2013) have recently claimed that gamification can be used effectively in the recruitment process to attract a large number of candidates, improve organizational image and attractiveness and, as a result, positively affect applicants' job pursuit behaviors toward an organization. Game elements might also improve the selection process, since it is more difficult for test-takers to fake the assessment, as desirable behaviors may be less obvious to individuals playing the game, and as a result, improve prediction of job performance and hiring decisions (Armstrong, Landers et al., 2016b). This could be especially the case for traditional selection methods, such as personality tests, which are prone to faking undermining thus their predictive validity (Murphy & Dzieweczynski, 2005). The gamification of selection methods is also likely to improve performance prediction by impeding information distortion and providing better quality information about the test-takers (Armstrong, Landers et al., 2016b). However, this is not an inherent quality of gamification but it largely depends on the type of gamification; candidates might be less likely to identify the correct or desirable answer and distort their answer, either intentionally to inflate their scores or unintentionally to be socially desirable (Richman, Kiesler, Weisband, & Drasgow, 1999).

Gamified assessment methods have also the potential to extract information about candidates' behavior more accurately compared to personality inventories (Armstrong, Landers et al., 2016b). Specifically, contrary to personality questionnaires, they do not include self-reported data. Instead of asking participants to indicate their agreement to various statements, they can assess gameplay behaviors to measure candidates' skills. These gameplay behaviors can be job related, and as a result, they might predict future work behavior more accurately than questionnaires (Armstrong, Landers et al., 2016b). Furthermore, Armstrong, Ferrell et al. (2016a), recently clarified that "a gamified assessment is not a stand-alone game, but it is instead an existing form of assessment that has been enhanced with the addition of game elements" (p. 672). This pertains that gamified assessments reflect an advanced level of existing typical types of selection methods, a meta-method that incrementally reinforces the possibilities of increased job performance prediction (Lievens, Peeters, & Schollaert, 2008).

Recently, different types of gamified assessment methods were developed by various specialized companies, such as Owiwi and Pymetrics, where others have focused on developing game-based assessments (e.g., Artic Shores and cut-e) attracting increased interest and use among organizations globally (Nikolaou, Georgiou, Bauer, & Truxillo, 2019). These gamified assessments might assess an applicant's cognitive ability or judgment regarding a situation encountered in the workplace. However, gamification types in employee selection vary and can include various elements, either narrative, such as additional text to an online questionnaire, till highly interactive game elements, such as avatars and digital rewards (Armstrong, Ferrell et al., 2016a). For example, gamified assessments might include virtual worlds sharing characteristics akin to work settings and avatars representing employees in order to assess candidates' skills and elicit job relevant behaviors (Laumer, Eckhardt, & Weitzel, 2012). Nevertheless, more research is needed to test the effectiveness of gamified assessment methods and establish valid and robust theoretical underpinnings to confirm their applicability in human resource management and employee selection settings.

On the other hand, research has already supported that SJTs predict job-related behaviors above cognitive ability and personality tests (Lievens et al., 2008). SJTs tend to determine behavioral tendencies, assessing how an individual will behave in a certain situation and are assumed to measure job and situational knowledge (Motowidlo, Dunnette, & Carter, 1990; Motowidlo, Hooper, & Jackson, 2006). Additionally, several scholars have concluded that SJTs can tap into a variety of constructs—ranging from problem solving and decision-making to interpersonal skills and they are able to measure multiple constructs at the same time (e.g., Christian, Edwards, & Bradley, 2010). Also recent research (Krumm et al., 2015; Lievens & Motowidlo, 2016) indicated that more general domain knowledge can be assessed by SJTs depending on the content of situations developed, leaving space to researchers and practitioners to better capture general soft skills' performance and increase the targeted audience administration.

Moreover, video technology has been successfully applied to SJTs increasing their effectiveness (e.g., Olson-Buchanan, Drasgow, Weekley, & Ployhart, 2006). To be more specific, the increased fidelity of presenting the situations in video format might lead to higher predictive validity whereas increased realism might result in favorable applicant reactions (Lievens & Sackett, 2006b). Oostrom, Born, Serlie, and Molen (2010) supported that an open-ended webcam SJT, utilizing a webcam instead of a static video recorder to capture the responses of participants, predicts job placement success. Rockstuhl, Ang, Ng, Lievens, and Dyne (2015) endeavored to

predict task performance and interpersonal OCB by expanding the traditional SJT paradigm to multimedia, implementing it across different cultural samples. In both cases, additional game elements in SJTs, such as a webcamera and video-based vignettes, respectively, contributed to better prediction of performance providing support in this practice as a promising method for personnel selection. More recently, Lievens (2017) suggested that webcam SJTs seem to be a promising approach in understanding intra-individual variability in controlled settings, by combining the procedural knowledge and the expressed behavior. It is also suggested that incorporating game elements into an existing practice in HR might have higher return on investment for an organization than developing a whole new digital game (Landers, 2014). Considering the psychometric qualities of SJTs (McDaniel, Hartman, Whetzel, & Grubb, 2007) along with the performance results when integrated with multimedia and game elements, the gamification of SJTs seems to be an appropriate method to follow. Armstrong, Ferrell et al. (2016a, p. 672), recently emphasized the role of gamification as "especially valuable to practitioners in an era moving toward business-to-consumer (B2C) assessment models" which is highly applicable for our research. Taking the above into consideration, we have chosen SJTs as the most appropriate methodology to develop and then, convert it into a new gamified assessment. To establish the effectiveness of the gamified selection method, we will initially explore the construct validity of a new SJT and the replication of the results to a gamified version of the test.

3 | GAMIFIED ASSESSMENT DEVELOPMENT

Our aim was to gamify an assessment method that would support organizations to map out prospective employees' soft skills. We first need to identify the most common core competencies and skills organizations often seek from their employees, especially when recruiting in graduate trainee and entry-level positions. For example, adaptability, flexibility, learning agility, knowledge breadth, and multicultural perspective have often been described as key competencies for employability across several stakeholder groups (e.g., Gray, 2016, "The digital future of work", 2017; Robles, 2012). Moreover, among the most common skills that individuals may use in several job positions are decision-making, flexibility, and ability to work under pressure, whereas on the other hand, employers often face difficulty in locating young graduates possessing soft skills, such as resilience and teamwork (Clarke, 2016). Following an extensive search of the literature and research on graduate employability, we selected four of the skills that seemed to become more and more relevant in today's demanding work environments (resilience, adaptability, flexibility, and decision-making) to form initially the SJT and subsequently the gamified assessment's dimensions.

We believe that these skills, which have been identified as key transferable soft skills integral to graduate employability (Andrews & Higson, 2008), are more suitable to be assessed through a gamified assessment than traditional selection methods, such as interviewing or psychometric tests. Moreover, many authors claim that the difficulty in transferring and assessing soft skills, compared to hard skills (e.g., technical or business-related knowledge and skills), results in increased waste of time and money for organizations (e.g., Laker & Powell, 2011), accounting for our focus on soft skills and need to use an assessment that may provide better quality information about candidates' behavior on the job. For example, Kubisiak, Stewart, Thornbury, and Moye (2014) employed self-report surveys to assess willingness to learn and a gamified simulation to assess ability to learn, concluding that a gamified assessment can be used to assess predictor constructs in a selection context, where survey methodology may not be adequate. Similarly, since resilience, ability, flexibility, and decision-making do not address intentions but behaviors, a gamified assessment might be better employed to measure these important attributes among job applicants.

Subsequently, we chose the type of gamification to employ. There is still limited research in human resources management and work/organizational psychology literature on gamification in selection and assessment but there are recommendations for researchers to approach gamified assessment addressing which game elements might affect and in what way the assessment outcomes (Armstrong, Ferrell et al., 2016a). Drawing from the taxonomy of gamification elements for use in educational contexts (Dicheva, Dichev, Agre, & Angelova, 2015), we gamified the SJT assessment in respect to the following gamification design principles: engagement, feedback, progress, freedom of choice, and storytelling. Although there are fundamental differences between game-based learning and gamified assessments, as the objective in learning is to motivate not to measure, the common gamification principles of game-based learning might also be appropriate for selection (Hawkes et al., 2017). Dicheva et al. (2015) reviewed previous studies on the application of gamification in education and mapped the context of application and game elements used. The game elements are conceptualized as the gamification design principles with the game mechanics that are typically used to implement them. For example, the game mechanisms that are used for the principles of engagement and feedback might be avatars (e.g., Deterding, Björk, Nacke, Dixon, & Lawley, 2013), immediate rewards instead of vague long-term benefits (e.g., Zichermann & Cunningham, 2011), and immediate or cycles of feedback (e.g., Nah, Zeng, Telaprolu, Ayyappa, & Eschenbrenner, 2014). In addition, the progress principle is achieved by using a progress bar or points and levels (e.g., Zichermann & Cunningham, 2011), while storytelling by using avatars (e.g., Nah et al., 2014), visual and voice overs. Finally, among the most common gamification design principles in educational settings, is freedom of choice (Dicheva et al., 2015), which in a gamified assessment, may relate to how players interact with the game as well as to other choices players may make, for example, whether they can skip a level, leave the game at any time, save it and return later, and so on (Hawkes et al., 2017).

We gamified the SJT assessment by using game mechanics that serve those principles. For example, in the beginning of the assessment, test-takers select a play hero/avatar. Every crew member appearing in the gamified assessment has a backstory. The story

follows the journey of play heroes in four islands, one for each soft skill assessed. Storytelling/narration takes place using visual and voice overs while playing the "game." We employed narration and fantasy in the gamified assessment to bring in engagement, meaning and clear calls to action, showing test-takers how to get on a path, in other words, to respond to scenarios. We could have used narration reflecting the real word but this could possibly have not the emotional advantages of fantasy and adventurous stories that keep people engaged. On the other hand, according to Malone and Lepper (1987), fantasy is described as one of the key reasons users appreciate in a game and as one of the most important features of games raising a player's imagination.

There is also a visual progression bar showing the progress in the assessment as well as story troubleshooting mechanisms and voice overs to remind users what the interface does and how to play the "game." There is also a world map showing the islands the players progress through. Rewards given to test-takers are intrinsic, by successfully completing the missions/solving the scenarios, and extrinsic, by receiving a report including feedback on player's competencies following the completion of the "game." Test takers have freedom of choice as well as they choose their avatar, they can skip the narrative and can leave the "game" anytime and continue from where they left off. Finally, a fine balance is kept between assessment and game mechanics to make it as fun and engaging as possible but without alienating nongamers from the experience and discriminate against them by making it fair and providing equal opportunities to all. In an adventure story setting, it might be more difficult to ascertain the context of the question making the candidate think twice and to respond with a more representative answer, while candidates' interest in the assessment might be increased.

4 | METHOD

4.1 | Samples

Initially, 20 experienced HR professionals in employee selection and assessment from various hierarchical levels (directors, managers, and recruiters), based in Athens, Greece, were interviewed during the development phase of the SJT. Also, seven HR professionals served as experts to determine the scoring key of the new SJT. For face validation purposes, another group of eight HR practitioners completed the SJT. Additionally, 321 business schools' students and graduates (61% female) with a mean age of 26.5 years old (*SD*: 5.4 years) and educational level: 42% bachelor's degree, 41% master's degree, were employed as the construct validity and confirmatory factor analysis sample. For conducting the replication of the gamified SJT, we gathered 410 employees or job seekers (46% female) on top of the 321 test takers of the previous step, with an average age of 29 years old (*SD*: 7.4 years), 72% of which had a bachelor's or master's degree.

4.2 | Measures

4.2.1 | SJT measurement

Twenty-five scenarios accompanied with four response options describing (a) Resilience, (b) Adaptability, (c) Flexibility, and (d) Decision-Making situations have been developed. Each scenario is accompanied by a scoring key, indicating the correct, wrong, and neutral alternatives. The participant should indicate which alternative serves as correct and wrong in each situation. Every correct choice gave +1 point to the test taker and −1 for the wrong choice, 0 points were given to the other two options. Each participant received four separate scores, one for each scale, which derives from summing up the individual scenario scores. A sample scenario of the SJT is presented in the Appendix.

In order to explore the construct validity of the SJT measure, assessing the four constructs, we used the following measures.

4.2.1.1 | Resilience

We used the Resilience Scale of Wagnild and Young (1993) which contains 25 items, all of which are measured on a 7-point scale from 1 (strongly disagree) to 7 (strongly agree). An example item is: "When I make plans I follow through with them." The alpha reliability of the scale was 0.89.

4.2.1.2 | Adaptability

We used the scale developed by Martin, Nejad, Colmar, and Liem (2012) consisting of nine items. Each item is measured on a 1 ("strongly disagree") to 7 ("strongly agree") scale. An example item is: "I am able to think through a number of possible options to assist me in a new situation." The alpha reliability of the scale was 0.89.

4.2.1.3 | Flexibility

Flexibility was measured using the HEXACO Personality Inventory (Lee & Ashton, 2004), which contains 10 items measured on a 5-point scale, from 1 ("strongly disagree") to 5 ("strongly agree") scale. An example item is: "I react strongly to criticism." The alpha reliability of the scale was 0.74.

4.2.1.4 | Decision-making

For the assessment of decision-making skills, researchers adopted Mincemoyer and Perkin's (2003) measure which assesses factors, such as "define the problem; generate alternatives; check risks and consequences of choices; select an alternative; and evaluate the decision." The response category for each question was a 5-point Likert-type scale (1 = never to 5 = always) designed to determine frequency of use. An example item is: "I easily identify my problem." The alpha reliability of the scale was 0.77.

4.3 | Procedure

We developed a SJT assessing the four competencies (resilience, adaptability, flexibility and decision-making) following the guidelines

suggested by Motowidlo et al. (1990). The content of SJT's situations and response options was first developed followed by an iterative procedure of face validation and construct validity assessment. At this stage, the SJT's scenarios along with measures of Resilience, Adaptability, Flexibility, and Decision-Making have been administered to business school students and graduates. Then, the SJT's scenarios were converted into adventure scenarios around a common story by a English-speaking professional writer. The professional writer converted the four competencies into "islands of adventure," then, authors thoroughly examined the content of the converted scenarios to ensure correspondence. A sample scenario of the gamified assessment is presented in the Appendix, where the players wonder around the islands responding to how they would most likely and least likely behave to particular instances. The mission of the gamified assessment is to respond to all situations/scenarios by indicating what is most likely and least likely to do. Having established a robust SJT measurement and a gamified equivalent, we proceeded to first ensure the construct validity of the gamified SJT and second to verify the lack of systematic variance between the two different modes of testing (SJT and gamified SJT). Therefore, we administered the gamified SJT to employees and job seekers for validation purposes. As a result, a fully functional gamified selection and assessment approach has been developed, which has been transferred in an online platform.

5 | STUDY 1: SJT DEVELOPMENT AND CONSTRUCT VALIDITY

5.1 | Item generation and content validation

Based on the critical incident methodology and experts' responses and also in retrospect with scenario's writing, four response options have been developed for each particular scenario, that elicit how the test taker will behave in each single situation (which took the form of "most likely would" and "least likely would"). We also employed subject matter experts' (SMEs) scoring approach (Bergman, Drasgow, Donovan, & Henning, 2006), which asks experts' opinion about which is the best and least likely response in each scenario.

Following the formal procedure of Haynes, Richard, and Kubany (1995), to further validate the content of the SJT, an empirical methodology called hit ratio analysis, initiated by Moore and Benbasat (1991), was performed. Experts indicated that six scenarios in Resilience, seven scenarios in Adaptability, six scenarios in Flexibility, and six scenarios in Decision-Making have survived after receiving recurrent refinements in terms of clarity and grammar. At the final stage, experts and researchers seem to reach acceptable levels of congruence after the procedure of content validation (ICC = 0.72, <0.01) in the 25-item SJT format, following the guidelines of Cicchetti (1994).

5.2 | Construct validity of the SJT

To ensure the results have not been influenced by Type II error, we performed a series of hierarchical linear regressions using the SJT facets as dependent variables, controlling for age and gender. The results presented in Table 1 provide evidence of convergent and discriminant validity of the SJT.

More specifically, the resilience SJT facet is related to the resilience scale at a significant but mediocre level (β = 0.350, p < 0.01), and so do the decision-making scale (β = 0.104, p < 0.05) and flexibility (β = −0.140, p < 0.05). On the other hand, the SJT flexibility measurement regression coefficients are statistically significant with the HEXACO personality inventory measuring flexibility (β = 0.366, p < 0.01) and adaptability scale (β = 0.166, p < 0.05). SJT Adaptability is related only to the adaptability scale (β = 0.166, p < 0.01) and the SJT Decision-Making facet with decision-making (β = 0.389, p < 0.01), flexibility (β = −0.114, p < 0.05) and resilience (β = 0.202, p < 0.01) scales, respectively. Some of the facets of SJT are cross-correlated with other measurements; however, the magnitude is low sufficing evidence for discriminant validity. To further establish convergent validity on the same sample (N = 321), we conducted CFA (Bentler, 2004) with maximum likelihood estimation and robust statistics to address nonnormality of data and fit indexes, as recommended by Hu and Bentler (1999). More specifically, a value of >0.90 for comparative fit index (CFI) and normed fit index (NFI) and a value of <0.05 for root mean square error of approximation (RMSEA) indicate a well-fitting model according to researchers. With the exception of the SJT Decision-Making and the respective scale, all other models present marginal though acceptable fit with the data (Table 2), satisfying thus to an extend the criteria for convergent validity.

The factor correlations in each specific model, ranging from 0.290 to 0.378 at a statistically significant level, provide evidence of convergence. Although fit is not strongly supported by these particular CFA models, RMSEA 90% CI at all cases is within the acceptable limits (MacCallum, Browne, & Sugawara, 1996), thus providing evidence for affirmative measurement even though chi-square and CFI are marginally accepted (Chen, Curran, Bollen, Kirby, & Paxton, 2008; Kenny & McCoach, 2003). Furthermore, according to Bagozzi, Yi, and Phillips (1991), CFA models can be utilized to address convergent and discriminant validity more effectively than Campbell and Fiske (1959) procedures and criteria and although model fit is one of the criteria to satisfy construct validity, it is not the most significant one. The reason is that some CFA criteria (i.e., χ^2) may be distorted due to small sample size and falsely neglect the actual correlation and covariance between the traits under investigation.

6 | STUDY 2: CONFIRMATORY FACTOR ANALYSIS AND REPLICATION STUDY

The platform hosting the gamified assessment has been released and 410 participants, voluntarily completed the on-line version of it (mainly employees or job seekers). To further establish the construct validity of the new, gamified version of the test, the authors picked a subsample of test-takers (mean age: 27,6, *SD*: 4,6) who had completed both the SJT and also played the gamified version of it

TABLE 1 Hierarchical linear regressions (N = 321) with SJT facets as dependent variables (Scales: independent variables)

Dependent variable: Resilience (SJT)						Dependent variable: Adaptability (SJT)					
	R^2	ΔR^2	*F* change	*B*	Sig		R^2	ΔR^2	*F* change	*B*	Sig
Step 1						*Step 1*					
Gender	0.215	0.215	3,498*	0.092	0.335	Gender	0.069	0.069	4,710	0.172	0.035*
Age				−0.120	0.117	Age				0.034	0.543
Step 2						*Step 2*					
Age	0.311	0.096	5,393*	−0.094	0.021*	Adaptability Scale	0.134	0.106	7,579	0.166	0.000**
Resilience Scale				0.350	0.000**						
Flexibility Scale				−0.140	0.041*						
Decision-Making Scale				0.104	0.045*						
Dependent variable: Flexibility (SJT)						**Dependent variable: Decision-Making (SJT)**					
	R^2	ΔR^2	*F* change	*B*	Sig		R^2	ΔR^2	*F* change	*B*	Sig
Step 1						*Step 1*					
Gender	0.087	0.087	1,144	0.069	0.227	Gender	0.067	0.067	1,154	0.032	0.458
Age				0.062	0.272	Age				0.016	0.670
Step 2						*Step 2*					
Flexibility Scale	0.281	0.106	7,157	0.366	0.000**	Decision-Making Scale	0.275	0.208	7,793	0.389	0.000**
Adaptability Scale				0.166	0.051	Resilience Scale				0.202	0.000**
						Flexibility Scale				−0.114	0.047*

Note. Table reports standardized beta coefficients; Resilience, Flexibility, Adaptability, and Decision-Making Scales as independent variables.
*Correlation is significant at the 0.05 level (two-tailed). **Correlation is significant at the 0.01 level (two-tailed).

TABLE 2 Confirmatory factor analysis results of the SJT (N = 321)

	S–B Scaled χ^2	CFI	NFI	RMSEA	RMSEA 90% CI	Resilience Scale	Adaptability Scale	Flexibility Scale	Decision-Making Scale	Model-fit
SJT resilience	817.7025 (p > 0.001)	0.890	0.870	0.050	0.042–0.058	0.352**				Accepted
SJT adaptability	205.7359 (p < 0.05)	0.854	0.843	0.058	0.046–0.070		0.378**			Marginal accepted
SJT flexibility	142.7897 (p > 0.001)	0.910	0.891	0.035	0.019–0.048			0.290**		Accepted
SJT Decision-Making	508.3967 (p < 0.05)	0.792	0.749	0.070	0.060–0.078				0.364**	Rejected

Note. **Correlation is significant at the 0.01 level (two-tailed).

(N = 97). Over this small sample, we performed linear regressions using as independent variables the set of well-established measurements and self-reports provided by the common subsample of test-takers. The results, after controlling for age and gender, revealed significant associations with the corresponding scales. Indicatively, the resilience facet in the gamified test is related to the measure of resilience (β = 0.565, p < 0.01), the adaptability facet to adaptability scale (β = 0.528, p < 0.01), and HEXACO scale (β = 0.187, p < 0.05), the flexibility facet to HEXACO flexibility and flexibility scale (β = 0.552, p < 0.01 and β = 0.211, p < 0.05, respectively) and the decision-making facet to the decision-making scale (β = 0.450, p < 0.01). Even though there are cross-loadings in some cases (i.e., flexibility and adaptability) their magnitude is small. It should be noted that the sample size is small and neither CFA or path analysis techniques are applicable due to potential identification errors (Kline, 2005).

To remedy this, a subsequent confirmatory factor analysis has been performed to both confirm the appropriateness of the test structure and get further insights into discriminant validation (N = 410). The results showed good fit to data (Satorra–Bentler Scaled χ^2 [269, N = 410] = 306.94, p = 0.05; CFI = 0.91; NNFI = 0.89; IFI = 0.91; RMSEA = 0.019; RMSEA 90% interval [0.000, 0.027]), with statistically significant coefficients' estimates ranging from 0.141 to 0.607 and zero covariances between dependent variables (i.e., constructs). This is an indication, according to Bagozzi et al. (1991), of discrimination between the facets, thus the resilience gamified facet does not covary with SJT flexibility, adaptability, and decision-making, adaptability is not significantly related with the resilience and decision-making SJT dimensions, flexibility presents no covariance with the resilience and decision-making SJT facets and the decision-making gamified dimension is not related to any of the other SJT facets. Additionally, residuals' analysis showed that residuals are symmetrically distributed around zero point, i.e, 95% of them is around zero (Joreskog & Sorbom, 1988), the average off-diagonal absolute standardized residuals are low, i.e., 0.04 (Bentler, 2004) and the Standardized Root Mean Square Residual (SRMR) of 0.48 is lower than the cut-off rate of 0.50 (Hu & Bentler, 1999). Moreover, the careful inspection of residual correlation matrix, proved no large magnitude of residual correlations ranging from −0.151 to 0.140, hence serving only to minimal discrepancy in fit variance between the hypothesized model and the sample data (Hu & Bentler, 1999). However, this marginal fit (p = 0.05) made the researchers suspicious of further model modifications to achieve a better fit and reassess unexplained model variance that may be detected to other equations' elements (Bentler, 2004).

To ensure that the transition from a paper and pencil SJT to a gamified environment was held smoothly, avoiding potential variance due to the utilization of different samples during the validation procedure, we employed cross-validation analysis over a joint sample of 321 university students, who took the SJT version, and 410 employees and job seekers, who took the gamified SJT version. Accordingly, multiple-group measurement invariance tests were performed on the SJT and gamified SJT scales, to assess cross-validation among samples. Previous research has shown that when parallel data exist

across groups, multiple-group analysis offers a powerful test of the equivalence of factor solutions across samples because it rigorously assesses measurement properties (Bagozzi & Yi, 1988; Bollen, 1989; Marsh, 1995; Marsh & Hocevar, 1985). Table 3 presents the fit estimates for the models in the invariance hierarchy.

The baseline model in all cases shows adequate fit with the data as all fit indices are within the predefined cut off values. With the exception of the fourth model, comparing with the baseline at all scale cases, all χ^2 differences along with the other fit indices indicate good fit to the data. Also, CAIC is decreasing after baseline model comparison, serving as an additional sign of nonchance lack of invariance. When the baseline model is constrained in factor loadings, factor correlations, and factor variances, all construct cases present large chi-square difference with the baseline models and adequate criteria values. More specifically, the chi-square difference from the baseline model is $\Delta\chi^2$ (13, N = 731) = 127.797, p < 0.001, in resilience, $\Delta\chi^2$ (15, N = 731) = 168.888, p < 0.001 in adaptability, $\Delta\chi^2$ (13, N = 731) = 230.504, p < 0.001 in flexibility, and $\Delta\chi^2$ (13, N = 731) = 195.547, p < 0.001 in decision-making, indicating large magnitude differences and hence rejection of the final hierarchy of models. Subsequently, the fit indices are lower than the cut-off values at all cases. These chi-square differences were relatively large; however, invariant factor variances are considered the least important in testing measurement property invariance across groups (Bollen, 1989; Marsh, 1995). Therefore, some evidence of partial measurement invariance is apparent across the samples (Vandenberg & Lance, 2000).

7 | DISCUSSION

The present study introduced a new gamified instrument to measure some of the skills and competencies that employers often look for when hiring young graduates. We gamified a SJT assessment measuring four constructs: resilience, adaptability, flexibility, and decision-making. These dimensions have been shown to be reliable and factorially distinct; whereas, the convergent and discriminant validity of the gamified measure was established by showing its associations with well-established self-report measures. To be more specific, having first developed and tested a SJT, we found preliminary support that the addition of game elements (e.g., avatars, feedback, narrative and visual/voice overs) to the SJT and its conversion into an adventure online story confirms the construct validity of the measure. Admitedly, the strength of the convergence described above is not well founded given that the majority of fit indexes are marginally acceptable by the existing literature (Hu & Bentler, 1999). However, as many scholars presumed in recent research, goodness of fit indexes are highly depending on sample sizes, estimators, or distributions and should be treated not as golden rules but supplementary to human judgment (Barrett, 2007; Bentler, 2007; Marsh, Hau, & Wen, 2004). For this reason, we tried to give ourselves some degrees of freedom in evaluating CFA models and base our decisions in more standardized evidence as provided by RMSEA index coupled with the accuracy of its estimation provided by RMSEA 90% Confidence Intervals. Indeed, a more thorough inspection of potential modifications in the item structure is needed, which should be a part of future research.

We believe that this novel instrument contributes to research and practice in two main ways. First, the current gamified assessment method is among the first validated instruments using game elements in order to assess candidates' soft skills. To the best of our knowledge, this is probably the first published study exploring the psychometric properties of a gamified selection method. Our study contributes to research on gamification and selection methods exploring the construct validity of a new gamified selection method emphasizing the use of gamification that focus on behavior and not traits. Contrary to personality inventories that include self-report data and are prone to social desirability bias (e.g., Mayer & Salovey, 1997; Roberts, Matthews, & Zeidner, 2010), a gamified assessment extracts information about candidates' intention and as a result might be less prone to faking and distortion. Accordingly, the scenarios incorporated in the gamified assessment are the SJT scenarios, which assess work-related behaviors and thus are more likely to predict future work behaviors than survey-based inventories (Armstrong, Ferrell et al., 2016a). The use of game elements might enable the test to assess skills indirectly, making it thus difficult for candidates to distort their answers, since the desirable behaviors are not so obvious to them. Also, a gamified assessment might enhance fun, motivation, and engagement as well as improve predictive validity (Collmus, Armstrong, & Landers, 2016; Yan, Conrad, Tourangeau, & Couper, 2010). Future research is needed to claim that the gamified assessment method we presented is more valid, fair, fun, or engaging beyond and above traditional selection methods, such as personality tests. We have taken the first steps to explore both applicant reactions, such as organizational attractiveness and recommendation intentions (Gkorezis, Georgiou, Nikolaou, & Perperidou, 2019; Nikolaou & Georgiou, 2017) and participants' performance (e.g., self-reported job and academic performance in Nikolaou, Georgiou, & Kotsasarlidou, 2019), providing preliminary support that the current gamified assessment has the potential to be an attractive and valid alternative to traditional selection methods.

Second, this study contributes to the literature on SJT. Our findings provide support that game elements, such as storylines, feedback, avatars, visual, and voice overs can be successfully applied to SJTs and effectively assess candidates' soft skills. Our study extends previous studies that dictated the applicability of SJTs in high fidelity modes, such as video, multimedia, and interactive formats (Lievens & Sackett, 2006a) and gamified contexts (Armstrong, Landers et al., 2016b). Webcams, static video recorders, and multimedia have been successfully applied to SJTs increasing the fidelity of presenting the scenarios leading to increased realism and more positive applicant reactions, as well as higher predictive validity (Lievens & Sackett, 2006b; Oostrom et al., 2010; Rockstuhl et al., 2015). The incorporation of games elements into a SJT is likely to increase fidelity, fun, fairness, and favorable applicant reactions, while eliciting behaviors and predicting job performance more successfully. Along these lines, our

TABLE 3 Cross-validation analysis results (N = 731)

Model	χ^2 (N = 731)	Df**	χ^2 diff*	Df χ^2 diff	RMSEA	GFI	CFI	CAIC	Model fit
Resilience									
No constraints (baseline)	62.104***	53	–	–	0.023 (0.000–0.044)	0.965	0.927	−295.61	ACCEPT
Factor loadings invariant	89.170***	63	27.066***	10	0.036 (0.016–0.053)	0.956	0.867	−336.441	ACCEPT
Factor loadings and factor correlation invariant	89.257***	64	27.153***	11	0.036 (0.015–0.052)	0.956	0.867	−342.704	ACCEPT
Factor loadings, factor correlation, and factor variances invariant	190.054***	66	127.797***	13	0.077 (0.065–0.090)	0.900	0.729	−255.405	REJECT
Adaptability									
No constraints (baseline)	106.830***	76	–	–	0.035				
(0.017–0.050)	0.956	0.895	−407.77	ACCEPT					
Factor loadings invariant	146.916***	88	40.086***	12	0.046 (0.032–0.058)	0.948	0.832	−448.42	ACCEPT
Factor loadings and factor correlation invariant	147.123***	89	40.207***	13	0.045 (0.032–0.058)	0.938	0.795	−454.97	ACCEPT
Factor loadings, factor correlation, and factor variances invariant	275.718***	91	168.888***	15	0.080 (0.069–0.091)	0.876	0.749	−339.91	REJECT
Flexibility									
No constraints (baseline)	61.707***	53	–	–	0.023 (0.000–0.044)	0.968	0.926	−296.34	ACCEPT
Factor loadings invariant	90.465***	63	28.758***	10	0.037 (0.018–0.053)	0.955	0.894	−335.14	ACCEPT
Factor loadings and factor correlation invariant	91.779***	64	30.072***	11	0.037 (0.018–0.053)	0.955	0.894	−340.58	ACCEPT
Factor loadings, factor correlation, and factor variances invariant	292.211***	66	230.504***	13	0.104 (0.092–0.116)	0.850	0.721	−153.66	REJECT
Decision-Making									
No constraints (baseline)	88.618***	53	–	–	0.046 (0.028–0.062)	0.946	0.897	−269.77	ACCEPT
Factor loadings invariant	120.854***	63	32.236***	10	0.054 (0.039–0.068)	0.942	0.827	−305.15	MARGINALLY ACCEPT
Factor loadings and factor correlation invariant	120.874***	64	32.256***	11	0.053 (0.038–0.067)	0.942	0.827	−311.89	MARGINALLY ACCEPT
Factor loadings, factor correlation, and factor variances invariant	284.165***	66	195.547***	13	0.102 (0.090–0.114)	0.850	0.716	−162.13	REJECT

Note. Empty cells indicate no calculation. The final 731 sample participants derived from the summation of 410 gamified SJT test takers and 321 SJT test takers. GFI, goodness-of-fit index; CFI, comparative fit index; CAIC, Conditional Akaike Information Criterion; diff., difference.

*Difference in the chi-square statistic between a given model and the baseline model. **Difference in degrees of freedom between a given model and the baseline model. ***p < 0.01.

study corresponds to the widespread use of the internet and game playing (Campbell, 2015) addressing calls for the exploration of the efficiency of gamification in employee selection.

7.1 | Practical implications

The current research has important practical implications for organizations. By establishing the construct validity of the gamified selection method, recruiters might use a new selection tool that effectively assesses soft skills, with the potential to reduce the risk and the "cost" of bad hires. Organizations might also improve their selection processes replacing or supplementing traditional selection methods with gamified selection methods.

Gamified selection methods share several benefits that other multimedia tests have. A gamified selection method can be administered over the Internet to a large group of applicants and on various locations while automatically recording candidates' responses (Oostrom, Born, & Van Der Molen, 2013). Also, it focuses on behavior and not on personality traits that appear to be a less important criterion in employee selection (Viswesvaran & Ones, 2000). Gamified selection methods might be used to obtain higher quality information from candidates since they are more difficult for test-takers to fake and better able to elicit behaviors than traditional selection methods (Armstrong, Landers et al., 2016b). Moreover, an interpersonally oriented multimedia SJT is expected to demonstrate higher criterion-related validity than a paper–pencil test (Lievens & Sackett, 2006a). The gamified SJT uses verbal and visional cues that enhance realism and as a result, might provide to future employers a superior assessment of candidates' skills compared to traditional selection tests. On top of that, these properties might positively affect applicants' reactions. Applicants perceive the multimedia tests as more valid and enjoyable and as a result, they are more satisfied with the selection process (Richman-Hirsch, Olson-Buchanan, & Drasgow, 2000).

Employers might also benefit from the use of a gamified selection method in increasing their organizational attractiveness and positive behavioral intentions, such as applicants' job offer acceptance ratio. Since organizations nowadays employ a diverse workforce, they should use a selection method that reduces adverse impact and respects ethnic minority (Oostrom et al., 2013). Hence, a multimedia-based assessment method is suggested to result in reduced adverse impact compared to a paper-and-pencil method.

7.2 | Limitations and future research

The present study is not without limitations. From a methodological point of view, the sample size is a main concern for the current study. As a matter of size, samples in these kind of studies should be larger. For example, the small sample size of 97 common test-takers of the paper and pencil and the gamified SJT approach is barely adequate (Hoyle, 1995; Kline, 2005) to perform robust statistical analyses and establish the construct validity of the gamified SJT. To this end, the results are not clear-cut and the conclusions are not easily interpretable. To remedy this, we performed a full confirmatory factor analysis with an independent larger sample (N = 410). The results of the CFA demonstrated a marginally acceptable fit for the three of the four scales of the game, with the exception of the decision-making scale, questioning thus the internal validity of the four-factors model. As a future endeavor, post hoc modifications are required which will lead to a potentially shorter version of our research "product," i.e., the gamified SJT, which will probably assist in achieving more robust results in forthcoming validation steps (criterion-related and incremental validation procedures). The size of the sample at the item generation stage is also a limiting factor in explaining our results. It would be better if we had employed a higher number of SMEs according to common practice (Bledow & Frese, 2009; Motowidlo et al., 2006).

In order to explore the construct validity of the gamified SJT, we reached a model of a marginal fit to data that may need further examination in the future, especially in criterion-related studies. Researchers proved that residuals are not an issue at this time and the marginal fit achieved is not in the scope of the current paper. However, further modification of the tested model should be performed following the guidelines described in the literature (e.g., Bentler, 2004), leading to a higher variance reallocation and therefore a different and probably lighter version of the gamified SJT.

From a practical viewpoint, this study is limited by the fact that the criterion-related or incremental validity of the assessment over traditional selection methods, such as personality or ability tests, has not been established yet (Nunnally & Bernstein, 1994), with the exception of self-reported job and academic performance (Nikolaou et al., 2019). Also, although we employed fantasy and adventurous stories that are likely to keep people engaged, there might be applicants who are in favor of games that are obviously related to the job. Future research could explore applicant reactions, such as perceived test fairness and appropriateness of selection instrument, among candidates who complete either or both the SJT and its gamified version to further support the effectivenss of using game elements into selection methods. Also, another limitation of the gamified SJT might be any accessibility issues for candidates who may not have the hardware or internet connection required to try the assement. Finally, the current version of the gamified assessment should also be enriched with the assessment of additional skills, since it currently only measures four skills.

7.3 | Conclusions

Recently, a number of organizations have employed the use of gamification and game-based assessments in employee recruitment and selection. However, no published empirical studies have explored the validity of gamification in assessing candidates' skills. Our study supports that converting a traditional SJT to a gamified assessment, in order to effectively assess candidates' soft skills, such as resilience, adaptability, and decision-making can be of value. We first presented the development of a SJT to form the basis of the gamified assessment method. Second, the SJT's construct validity was

explored in order to transform it into a gamified assessment method, and third, the construct validity of the gamified assessment was established. As a result, the current study contributes to research on the use of game elements in employee selection as well as to SJT research and practice. By eliciting job-relevant behaviors within the context of a gamified assessment, increased prediction of future work behaviors may be possible compared to traditional psychometric tests. Future research is needed in order to provide further support that gamified assessment methods can be an accurate and attractive selection method.

ORCID

Konstantina Georgiou https://orcid.org/0000-0001-8272-9089
Ioannis Nikolaou https://orcid.org/0000-0003-3967-5040

REFERENCES

Andrews, J., & Higson, H. (2008). Graduate employability, 'soft skills' versus 'hard' business knowledge: A European study. *Higher Education in Europe*, *33*(4), 411–422. https://doi.org/10.1080/03797720802522627

Armstrong, M. B., Ferrell, J., Collmus, A., & Landers, R. (2016a). Correcting misconceptions about gamification of assessment: More than SJTs and badges. *Industrial and Organizational Psychology*, *9*(3), 671–677.

Armstrong, M. B., Landers, R. N., & Collmus, A. B. (2016b). Gamifying recruitment, selection, training, and performance management: Game-thinking in human resource management. In D. Davis & H. Gangadharbatla (Eds.), *Emerging research and trends in gamification* (pp. 140–165). Hershey, PA: IGI Global.

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *16*, 74–94. https://doi.org/10.1007/BF02723327

Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing construct validity in organizational research. *Administrative Science Quarterly*, *36*, 421–458. https://doi.org/10.2307/2393203

Barrett, P. (2007). Structural equation modeling: Adjusting model fit. *Personality and Individual Differences*, *42*, 815–824.

Bentler, P. (2004). *EQS 6 structural equations program manual*. Encino, CA: Multivariate Software.

Bentler, P. (2007). On tests and indices for evaluating structural models. *Personality and Individual Differences*, *42*, 825–829. https://doi.org/10.1016/j.paid.2006.09.024

Bergman, M. E., Drasgow, F., Donovan, M. A., Henning, J. B., & Juraska, S. E. (2006). Scoring situational judgment tests: Once you get the data, your troubles begin. *International Journal of Selection and Assessment*, *14*(3), 223–235. https://doi.org/10.1111/j.1468-2389.2006.00345.x

Bledow, R., & Frese, M. (2009). A situational judgment test of personal initiative and its relationship to performance. *Personnel Psychology*, *62*, 229–258. https://doi.org/10.1111/j.1744-6570.2009.01137.x

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

Campbell, C. (2015). *Here's how many people are playing games in America*. Retrieved from http://www.polygon.com/2015/4/14/8415611/gaming-stats-2015

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. https://doi.org/10.1037/h0046016

Chamorro-Premuzic, T., Akhtar, R., Winsborough, D., & 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

Chen, F., Curran, R. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). Evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological Methods & Research*, *36*(4), 462–494.

Chow, S., & Chapman, D. (2013). *Gamifying the employee recruitment process*. Paper presented at the Proceedings of the First International Conference on Gameful Design, Research, and Applications.

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

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Clarke, M. (2016). Addressing the soft skills crisis. *Strategic HR Review*, *15*(3), 137–139. https://doi.org/10.1108/SHR-03-2016-0026

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* (pp. 103–124). Cham, Switzerland: Springer International.

Deterding, S., Björk, S. L., Nacke, L. E., Dixon, D., & Lawley, E. (2013). *Designing gamification: creating gameful and playful experiences*. Paper presented at the CHI'13 Extended Abstracts on Human Factors in Computing Systems.

Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). Gamification in education: A systematic mapping study. *Journal of Educational Technology & Society*, *18*(3), 75.

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.), *The Wiley Blackwell handbook of the psychology of recruitment, selection and employee retention* (pp. 293–309). Chichester, UK: John Wiley & Sons.

Gkorezis, P., Georgiou, K., Nikolaou, I., & Perperidou, S. (2019). *Game-based assessment vs situational judgment test: Applicant reactions and recruitment outcomes through a moderated mediation model*. Paper accepted for presentation to the 19th EAWOP Congress, Turin, Italy.

Gray, A. (2016, January 19). *The 10 skills you need to thrive in the fourth industrial revolution*. Retrieved from https://www.weforum.org/agenda/2016/01/the-10-skills-you-need-to-thrive-in-the-fourth-industrial-revolution

Hawkes, B., Cek, I., & Handler, C. (2017). The gamification of employee selection tools: An exploration of viability, utility, and future directions. In J. C. Scott, D. Bartram, & D. H. Reynolds (Eds.), *Next generation technology-enhanced assessment: Global perspectives on occupational and workplace testing* (pp. 288–313). New York, NY: Cambridge University Press.

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment. A functional approach to concepts and methods. *Psychological Assessment*, *7*, 238–247. https://doi.org/10.1037/1040-3590.7.3.238

Hoyle, R. (1995). *Structural equation modeling: Concepts, issues, and applications*. Thousand Oaks, CA: Sage.

Hu, L., & Bentler, P. (1999). Cut off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55. https://doi.org/10.1080/10705519909540118

Joreskog, K. G., & Sorbom, D. (1988). *Prelis-A program for multivariate data screening and data summarization. A preprocessor for LISREL*. Chicago, IL: Scientific Software.

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*(3), 333–351. https://doi.org/10.1207/S15328007SEM1003_1

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: The Guilford Press.

Krumm, S., Lievens, F., Hüffmeier, J., Lipnevich, A. A., Bendels, H., & Hertel, G. (2015). How "situational" is judgment in situational judgment tests? *Journal of Applied Psychology*, *100*(2), 399–416.

Kubisiak, C., Stewart, R., Thornbury, E., & Moye, N. (2014, May). Development of PDRI's learning agility simulation. In E. C. Popp (Chair), *Challenges and innovations of using game-like assessments in selection*. Symposium presented at the 29th Annual Conference of the Society for Industrial and Organizational Psychology, Honolulu, HI.

Laker, D. R., & Powell, J. L. (2011). The differences between hard and soft skills and their relative impact on training transfer. *Human Resource Development Quarterly*, *22*(1), 111–122. https://doi.org/10.1002/hrdq.20063

Landers, R. N. (2014). Developing a theory of gamified learning: Linking serious games and gamification of learning. *Simulation & Gaming*, *45*(6), 752–768. https://doi.org/10.1177/1046878114563660

Laumer, S., Eckhardt, A., & Weitzel, T. (2012). Online gaming to find a new job-examining job seekers' intention to use serious games as a self-assessment tool. *German Journal of Human Resource Management*, *26*(3), 218–240. https://doi.org/10.1177/239700221202600302

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, *39*(2), 329–358. https://doi.org/10.1207/s15327906mbr3902_8

Lievens, F. (2017). Integrating situational judgment tests and assessment centre exercises into personality research: Challenges and further opportunities. *European Journal of Personality*, *31*(5), 487–502.

Lievens, F., & De Soete, B. (2012). Simulations. In S. Schmitt (Ed.), *The Oxford handbook of personnel assessment and selection* (pp. 383–410). New York, NY: Oxford University Press.

Lievens, F., & Motowidlo, S. J. (2016). Situational judgment tests: From measures of situational judgment to measures of general domain knowledge. *Industrial and Organizational Psychology*, *9*(1), 3–22. https://doi.org/10.1017/iop.2015.71

Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: A review of recent research. *Personnel Review*, *37*(4), 426–441. https://doi.org/10.1108/00483480810877598

Lievens, F., & Sackett, P. R. (2006a). Situational judgment tests in high stakes settings: Issues and strategies with generating alternate forms. *Journal of Applied Psychology*, *92*(4), 1043–1055.

Lievens, F., & Sackett, P. R. (2006b). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology*, *91*(5), 1181.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*, 130–149. https://doi.org/10.1037/1082-989X.1.2.130

Malone, T. W., & Lepper, M. (1987). *Making learning fun: A taxonomy of intrinsic motivations for learning*. Hills-Dale, NJ: Erlbaum.

Marsh, H. W. (1995). Confirmatory factor analysis models of factorial invariance: A multifaceted approach. *Structural Equation Modeling*, *1*, 5–34. https://doi.org/10.1080/10705519409539960

Marsh, H. W., Hau, K., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's 1999 findings. *Structural Equation Modeling*, *11*, 320–341. https://doi.org/10.1207/s15328007sem1103_2

Marsh, H. W., & Hocevar, D. P. (1985). The application of confirmatory factor analysis to the study of self-concept: First and higher order factor structures and their invariance across age groups. *Psychological Bulletin*, *97*, 562–582.

Martin, A. J., Nejad, H., Colmar, S., & Liem, G. A. (2012). Adaptability: Conceptual and empirical perspectives on responses to change, novelty and uncertainty. *Australian Journal of Guidance and Counselling*, *22*, 58–81. https://doi.org/10.1017/jgc.2012.8

Mayer, J.-S., & Salovey, A. (1997). What is emotional intelligence. In P. S. D. J. Salovey (Ed.), *Emotional development and emotional intelligence: Implications for educators* (pp. 3–31). New York, NY: Harper Collins.

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

Michael, D. R., & Chen, S. L. (2005). *Serious games: Games that educate, train, and inform*. Boston, MA: Muska & Lipman/Premier-Trade.

Mincemoyer, C. C., & Perkins, D. F. (2003). *Assessing decision-making skills of youth*. Paper presented at the Forum for Family and Consumer Issues.

Moore, G. C., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting and information technology innovation. *Information Systems Research*, *2*(3), 192–222.

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology*, *75*(6), 640–647. https://doi.org/10.1037/0021-9010.75.6.640

Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006). Implicit policies about relations between personality traits and behavioural effectiveness in situational judgement items. *Journal of Applied Psychology*, *91*, 749–761.

Murphy, K. R., & Dzieweczynski, J. L. (2005). Why don't measures of broad dimensions of personality perform better as predictors of job performance? *Human Performance*, *18*(4), 343–357.

Nah, F.-F.-H., Zeng, Q., Telaprolu, V. R., Ayyappa, A. P., & Eschenbrenner, B. (2014). *Gamification of education: A review of literature*. Paper presented at the International conference on hci in business.

Nikolaou, I., & Georgiou, K. (2017). *Serious gaming and applicants' reactions; The role of openness to experience*. In M. Armstrong, D. R. Sanchez & K. N. Bauer (2017): Gaming and gamification IGNITE: Current trends in research and application. 32nd Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.

Nikolaou, I., Georgiou, K., Bauer, T. N., & Truxillo, D. M. (2019). Technology and applicant reactions. In R. N. Landers (Ed.), *Cambridge handbook of technology and employee behavior* (pp. 100–130). Cambridge, UK: Cambridge University Press.

Nikolaou, I., Georgiou, K., & Kotsasarlidou, V. (2019). Exploring the relationship of a game-based assessment with performance. *The Spanish Journal of Psychology*, *22*, E6.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York, NY: McGraw-Hill Inc.

Olson-Buchanan, J. B., & Drasgow, F. (2006). Multimedia situational judgment tests: The medium creates the message. In J. Weekley & R. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 253–278). San Francisco, CA: Jossey-Bass.

Oostrom, J. K., Born, M. P., Serlie, A. W., & Van Der Molen, H. T. (2010). Webcam testing: Validation of an innovative open-ended multimedia test. *European Journal of Work and Organizational Psychology*, *19*, 532–550. https://doi.org/10.1080/13594320903000005

Oostrom, J. K., Born, M. P., & van der Molen, H. T. (2013). Webcam tests in personnel selection. *The Psychology of Digital Media at Work*, 166–180.

Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and

interviews. *Journal of Applied Psychology*, *84*(5), 754–775. https://doi.org/10.1037/0021-9010.84.5.754

Richman-Hirsch, W. L., Olson-Buchanan, J. B., & Drasgow, F. (2000). Examining the impact of administration medium on examinee perceptions and attitudes. *Journal of Applied Psychology*, *85*(6), 880–887. https://doi.org/10.1037/0021-9010.85.6.880

Roberts, R. D., Matthews, G., & Zeidner, M. (2010). Emotional intelligence: Muddling through theory and measurement. *Industrial and Organizational Psychology*, *3*(2), 140–144. https://doi.org/10.1111/j.1754-9434.2010.01214.x

Robles, M. M. (2012). Executive perceptions of the top 10 soft skills needed in today's workplace. *Business Communication Quarterly*, *75*(4), 453–465. https://doi.org/10.1177/1080569912460400

Rockstuhl, T., Ang, S., Ng, K. Y., Lievens, F., & Van Dyne, L. (2015). Putting judging situations into situational judgment tests: Evidence from intercultural multimedia SJTs. *Journal of Applied Psychology*, *100*, 464–480. https://doi.org/10.1037/a0038098

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*(2), 262–274. https://doi.org/10.1037/0033-2909.124.2.262

The digital future of work: What skills will be needed? (2017, July). Retrieved from https://www.mckinsey.com/global-themes/future-of-organizations-and-work/the-digital-future-of-work-what-skills-will-be-needed

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70. https://doi.org/10.1177/109442810031002

Viswesvaran, C., & Ones, D. S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment*, *8*(4), 216–226. https://doi.org/10.1111/1468-2389.00151

Wagnild, G. M., & Young, H. M. (1993). Development and psychometric evaluation of the resilience scale. *Journal of Nursing Management*, *1*(2), 165–178.

Yan, T., Conrad, F. G., Tourangeau, R., & Couper, M. P. (2010). Should I stay or should I go: The effects of progress feedback, promised task duration, and length of questionnaire on completing web surveys. *International Journal of Public Opinion Research*, *23*(2), 131–147. https://doi.org/10.1093/ijpor/edq046

Zichermann, G., & Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in web and mobile apps*. Cambridge, MA: O'Reilly Media Inc.

SUPPORTING INFORMATION

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

How to cite this article: Georgiou K, Gouras A, Nikolaou I. Gamification in employee selection: The development of a gamified assessment. *Int J Select Assess*. 2019;00:1–13. https://doi.org/10.1111/ijsa.12240