



Journal of Work and Organizational Psychology

<https://journals.copmadrid.org/jwop>



Predictive Validity, Applicant Reactions, and Influence of Personal Characteristics of a Gamefully Designed Assessment

Pedro J. Ramos-Villagrasa and Elena Fernández-del-Río

Universidad de Zaragoza, Spain

ARTICLE INFO

Article history:

Received 23 April 2023
Accepted 9 October 2023

Keywords:

Serious games
Game-related assessment
Game-based assessment
Personnel selection
Academic performance
Job performance
Personality

Palabras clave:

Juegos serios
Evaluación relacionada con juegos
Evaluación basada en juegos
Selección de personal
Rendimiento académico
Desempeño laboral
Personalidad

ABSTRACT

The use of game-related assessments (GRAs) in personnel selection is increasing. However, the type of GRA under investigation should be considered. This study focuses on game-based assessments, which closely resemble leisure-oriented videogames. We specifically examined a serious game that assesses personality traits (assertiveness, sociability, tolerance, and rules). We analyzed its predictive validity, applicant reactions, and the impact of personal and job-related factors on scores. The study included 182 participants who played the GRA and completed a Big Five questionnaire. Findings revealed that the game predicts adaptive performance, offering incremental value over the Big Five ($\Delta R^2 = .107$). Additionally, candidates reacted more positively to the GRA in terms of comfort and fit, although the effect size was small ($d = -0.1757$ and -0.3119 , respectively). Relating to personal and job-related characteristics, only job experience significantly affected the assertiveness score, albeit with a small effect size ($d = -0.3870$).

Validez predictiva, reacciones de los candidatos e influencia de las características personales de una evaluación que utiliza juegos

RESUMEN

Aumenta en el campo de la selección de personal la utilización de la evaluación relacionada con juegos (ERJ). El estudio se centra en la evaluación mediante juegos, un tipo de ERJ cercano a los videojuegos recreativos convencionales. En concreto se analizó un juego serio que mide rasgos de personalidad (asertividad, sociabilidad, tolerancia y reglas), investigando su validez predictiva, las reacciones de los candidatos y la influencia de las variables personales y laborales. En la investigación participaron 182 personas que jugaron a la ERJ y cumplimentaron un cuestionario de personalidad Big Five. Los resultados muestran la validez incremental de la ERJ con respecto a los Big Five en la predicción del desempeño adaptativo ($\Delta R^2 = .107$). Además, las reacciones fueron más positivas a la comodidad y al ajuste al puesto para la ERJ que para la prueba de personalidad, aunque con un tamaño del efecto modesto ($d = -0.1757$ y -0.3119). Con respecto a las características personales y laborales, la experiencia laboral se asoció a una mayor asertividad, aunque el efecto también es pequeño ($d = -0.3870$).

The rise of digital personnel selection procedures has led to a growing interest in new assessment methods (Andrés et al., 2023), such as digital interviews and testing (Woods et al., 2020). Among these approaches, game-related assessments (GRAs) have recently attracted significant attention (Landers & Sanchez, 2022). Their popularity is mainly based on the idea that they reduce the risk of faking and improve applicants' reactions while maintaining predictive validity (Melchers & Basch, 2022; Wu et al., 2022). However, research on GRAs invites skepticism and seeking more evidence (Ramos-Villagrasa et al., 2022). This study extends prior literature by focusing on a type of, until now, scarcely researched type of GRA (i.e., game-based assessments), considering four

job performance dimensions (i.e., task performance, contextual performance, counterproductive work behaviors, and adaptive performance). It also analyzes applicants' reactions to different issues (i.e., perception of comfort and suitability) and investigates the influence of individual differences in scores and reactions to game-based assessments.

GRAs: Types and Characteristics

GRAs refer to the application of gamification science to assessment contexts. Gamification science allows the use of game elements in non-gaming contexts such as personnel selection or training (Landers

Cite this article as: Ramos-Villagrasa, P. J., & Fernández-del-Río, E. (2023). Predictive validity, applicant reactions, and influence of personal characteristics of a gamefully designed assessment. *Journal of Work and Organizational Psychology*, 39(3), 169-178. <https://doi.org/10.5093/jwop2023a18>

Funding: This work was supported by the Ministry of Science and Innovation, Government of Spain, under grant PID2021- 122867NA-I00; and the Government of Aragon (Group S31_20D), Department of Innovation, Research and University. Correspondence: pjramos@unizar.es (P. J. Ramos-Villagrasa).

ISSN: 1576-5962/© 2023 Colegio Oficial de la Psicología de Madrid. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

	Purpose	Type	Characteristics
Degree of Playfulness ↓	Serious (assessment)	Soft-gamified assessments	The assessment is conducted in a similar way like traditional assessments (e.g., Likert scales, forced-choice scales, situational judgment test). Measures theoretically-driven constructs. Includes one or more game elements (e.g., story, music, game scores). If game elements were removed, the assessment could still be performed.
		Hard-gamified assessments (also “gamefully designed assessments”)	Like soft-gamified assessments, but game elements are embedded in the assessment (e.g., the story is necessary to understand the item; the control of the avatar is necessary to respond to the item). Consequently, game elements cannot be removed from the assessment. Measures theoretically-driven constructs.
		Game-based assessments	The way in which the assessment is performed is mainly like play a game. The behavior in the game (e.g., mouse clicks, reaction time) is used to develop the game scores. The constructs measured can be theory-driven or data-driven. Complementarily, the game can included elements from traditional assessments.
	Entertaining	Playful games	The behavior of the player (e.g., mouse clicks, reaction time) and/or the outcomes of the game (e.g., scores, level reached in a single round). Measures data-driven constructs.

Figure 1. Types of Game-related Assessments (GRAs).

et al., 2018). Focusing on personnel selection, the use of GRAs in real contexts has increased in the last few years for at least two reasons. Firstly, due to the digitalization of Human Resources Management (Nikolaou, 2021) as GRAs are mainly developed through technology (Melchers & Basch, 2022). Secondly, because GRAs seem to be a way to solve typical issues in personnel selection like faking or negative applicant reactions (Landers & Sanchez, 2022).

Nonetheless, research on this topic is growing and providing a less optimistic perspective. The recent review of the literature performed by Ramos-Villagrasa et al. (2022) summarized the existing evidence with respect to GRAs. According to them, it appears that GRAs are a promising selection method, but the accumulated evidence is still scarce and yields mixed results. Focusing on the inconclusive results regarding GRAs, they suggest that differences may be due to the particularities of each type of GRA. Considering this, research should take into account which type of GRA is under study and avoid generalizing research outcomes from one type to another (Ohlms et al., 2023).

According to Ramos-Villagrasa et al. (2022), differences among GRAs are based on their degree of playfulness (the degree in which it is close to traditional assessment or to games designed for entertainment) and their purpose (whether is a serious game designed for assessment purposes, or if it is designed for fun and is being used for assessment). Figure 1 outlines this classification and the main characteristics of the different types of GRAs: soft-gamified assessments, hard-gamified assessments, game-based assessments, and playful games. Each type is described below.

Soft-gamified assessments incorporate elements from games to enhance the assessment experience, while retaining the fundamentals of traditional assessment methods. It is used to measure the theoretically supported constructs (e.g., cognitive ability, personality, competencies). The assessment approach remains traditional in terms of evaluation (e.g., the use of Likert scales). It is considered a soft gamification process to develop the GRA because if game elements are removed, the assessment can still be performed. An example of this type of assessment is the gamified version of the Wisconsin Card Sorting Test (Hommel et al., 2022).

The second type of GRAs is hard-gamified assessments, also known as gamefully designed assessments (Ramos-Villagrasa et al., 2022). This type shares many characteristics with the soft one but is considered a “hard” version because the gamification process is deeply integrated into the assessment. Consequently, measuring the constructs of interest is impossible without the gamified elements. Similar to the soft type, it measures theoretically driven constructs. An example is Owiwi (Nikolaou et al., 2019), a situational judgment test in which item content can be understood only within the context of the game’s narrative.

The third type of GRAs related to serious purposes is game-based assessments. This type is closely related to playful games that evaluate the constructs of interest using users’ in-game behavior (e.g., reaction time, mouse clicks) instead of, or in addition to, traditional assessment methods such as answering questions or providing solutions to a set of predefined options. The constructs measured with game-based assessments can be theory-driven or data-driven. If they are data-driven, demonstrating the construct-related validity or predictive validity of the constructs measured by the game is needed. An example of this type of GRAs is Cognify (Auer et al., 2022).

The last type of GRAs is playful games. Playful games are designed for entertainment but they are repurposed for assessment. Consequently, any kind of construct-related validity or predictive validity study is needed. An example is the virtual reality game Job Simulator (Simons et al., 2023).

Ramos-Villagrasa et al. (2022) use this classification to clarify mixed results of prior research. They suggest that GRAs closely resembling traditional assessments (i.e., soft- and hard-gamified assessments) demonstrate better psychometric characteristics (reliability, construct validity, and predictive validity), whereas those close to actual games (i.e., game-based assessments and playful games) elicit more favorable applicant reactions. The only exception is the relationship with job performance, although this can be mitigated by providing explanations to the individuals being evaluated on this matter (Georgiou, 2021).

In any event, as the review itself suggests, further research is needed to fill the existing gaps for each type of GRA. Following this idea, we focus on serious games that are closer to playful games: game-based assessments.

Game-based Assessments

To the best of our knowledge, only five studies have provided empirical evidence of the use of game-based assessments (Ramos-Villagrasa et al., 2022). The GRAs used in these studies measure cognitive ability (Auer et al., 2022; Landers et al., 2021), competencies (Albadán et al., 2016; Wiernik et al., 2022), or personality (Wu et al., 2022). To summarize prior research, we are going to focus on construct validity, predictive validity, applicant reactions, and personal characteristics.

Concerning construct validity, Auer et al. (2022) investigated if trace data modelling may improve the prediction of cognitive ability of the game-based assessment Cognify and if this approach also allows to measure conscientiousness. Their study, performed with undergraduate students, found support only for cognitive ability. Similarly, Albadán et al. (2016) suggest the use of fuzzy logic

to improve applicant scoring. Another study, performed by [Wu et al. \(2022\)](#), used two game-based assessments (i.e., Click Town and Word Find) to measure the Big Five, but they found that these games actually measure cognitive ability better than personality. Another relevant issue is discriminant validity: [Wiernik et al. \(2022\)](#) provided evidence that their game-based assessment developed for the U.S. Air Force, Virus Slayer, measures different competences relevant to cybersecurity positions. Overall, the results suggest that game-based assessment has more difficulties in providing construct validity than other types of GRAs, such as gamified assessments and gamefully designed assessments (e.g., [Georgiou et al., 2019](#); [Landers & Collmus, 2022](#)). To solve this issue, [Auer et al. \(2022\)](#) highlighted the importance of defining the constructs that are going to be measured at the beginning of game design.

As for predictive validity, game-based assessments have demonstrated their ability to predict academic performance ([Auer et al., 2022](#); [Landers et al., 2021](#)) and task performance ([Landers et al., 2021](#)). However, job performance is a multidimensional construct and other dimensions, namely contextual performance, counterproductive work behaviors, and adaptive performance, may be considered ([Ramos-Villagrasa et al., 2019](#)). The more the dimensions of performance a GRA can predict, the greater its versatility for different organizations and job positions.

As far as applicant reactions are concerned, only one study has addressed this issue using game-based assessments: [Landers et al. \(2021\)](#). They compare applicants' reactions to a set of game-based assessments (i.e., Numbubbles, Resemble, Grid Lock, Proof It, Tally Up, Colour Pop, and Short Cuts) and to traditional assessment. They found better results for the games than for the traditional assessment, but the size of the improvement was low. This outcome is less optimistic than other studies that use other GRAs (e.g., [Harman & Brown, 2022](#); [Hommel et al., 2022](#); [Landers & Collmus, 2022](#)), encouraging further research to be sure whether this is a characteristic of game-based assessments in general or of the games used by Landers et al.

Personal characteristics are another issue of interest in GRAs, because any variable that may impact on GRAs scores should be controlled. In that sense, [Wiernik et al. \(2022\)](#) found a positive effect of education on Virus Slayer scores. However, [Landers et al. \(2021\)](#) did not find any evidence of gender differences. Research on other types of GRAs has investigated the influence of other characteristics such as age or the use of technology and videogames. Their findings suggest that the influence of these characteristics is modest ([Melchers & Basch, 2022](#)). Nevertheless, empirical research must check if this is the case for game-based assessments.

Taking the results as a whole, we may draw some conclusions: (1) the number of studies on game-based assessments is still scarce, then more primary research is needed to further investigate of differences between types of GRAs; (2) game-based assessments should be based on existing psychological theories and constructs to ensure construct validity; (3) the predictive validity of game-based assessments can be improved by investigating their ability to predict other dimensions of job performance; (4) we still know little about applicant reactions to game-based assessments; and (5) the influence of certain personal characteristics like age, the use of technology or experience with videogames should be investigated using game-based assessments. The aim of the present study is to advance research on GRAs in general and game-based assessments in particular.

The Present Study

The game-based assessment investigated in this study was Nawaiam (<https://nawaiam.com>), a commercial GRA used for selection, career counseling, and team management purposes. The study was conducted in collaboration with the company that owns the game, which provided access codes for Nawaiam to students and

alumni from the authors' university. The company did not influence the development or results of the study.

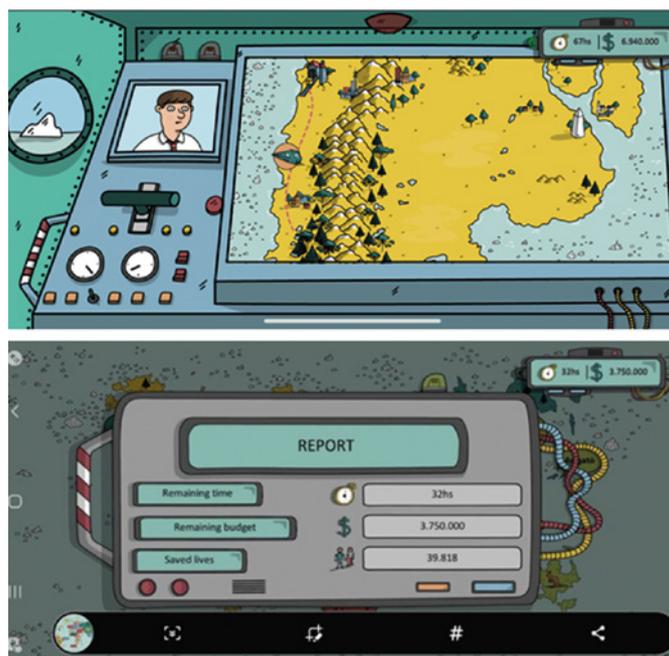


Figure 2. Screenshots of Nawaiam Game-based Assessment.

Nawaiam is set in the near future where the polar ice caps have melted. The player must make difficult decisions that contribute to the survival of the human race (related to managing resources, helping people, etc.), and must also overcome some skill-based games that are not part of the evaluation (i.e., they are involved in the assessment, but as an additional part of the gaming experience). Sample images are shown in [Figure 2](#). A description of the game following the taxonomy of [Bedwell et al.'s \(2012\)](#) game elements is shown in [Table 1](#). Nawaiam is different from other game-based assessments because it emphasizes the environment, game fiction, and immersion.

Nawaiam is played on the applicants' mobile phone and lasts around 20 minutes. Subsequently, the applicant is described in terms of one of the game's several "behavioral profiles" (i.e., behavioral dispositions such as personality) inspired by the DISC personality model ([Marston, 1928](#)). According to [Rodríguez \(2004\)](#), the DISC personality model describes people according to four dimensions based on two axes: a friendly/unfriendly relationship with the environment and an active/passive behavior. Considering both axes, the dimensions are Dominance (D, unfriendly/active), which characterizes people with confidence and focused on achieving outcomes; Influence (I, friendly/active), which describes people prone to persuade and cultivate connections with others; Steadiness (S, friendly/passive), which is associated with disposition to cooperation and being honest with others; and Conscientiousness (C, unfriendly/passive), which emphasizes interest in quality and accuracy. Scales tend to measure DISC traits using forced-choice items. [Owen et al. \(2020\)](#) state that the DISC model has construct validity, reliability, and may be used in multiple industries, sectors, and cultures. Nevertheless, although the DISC test has been administered to more than 50 million people in professional practice, empirical research studies using this test are really limited ([Gómez et al., 2021](#)). Due to the small amount of studies employing the DISC, its association with other personality models, such as the Big Five, remains unclear. Some studies connect negative emotionality with D and I; extraversion and open-mindedness with all DISC traits; agreeableness with D, S, and C; and conscientiousness with C ([Gehrig & Bonnstetter, 2017](#); [Jones](#)

Table 1. Description of the Assessment according to Bedwell et al.'s (2012) Taxonomy

Category	Description
1. Action language	The player must touch their mobile phone's screen to choose options or affect the environment.
2. Assessment	The player's decisions score on one of the four dimensions used to develop the behavioral profile (assertiveness, sociability, tolerance, and rules). The player does not have information about their scores throughout the game.
3. Conflict/ Challenge	The story is linear but with changes depending on players' decisions. Difficulty is low because any decision makes the story go on.
4. Control	The play choose between different actions at certain points of the story.
5. Environment	The story takes place in an apocalyptic version of our world in the near future.
6. Game fiction	High degree of realism in the narrative and course of action, but using pleasant pictures.
7. Human interaction	There is no human interaction throughout the game.
8. Immersion	The game uses pictures, video, and music to enhance immersion.
9. Rules/Goals	The rules are clear and known to the player at the start of the game.

& Hartley, 2013). Hence, it seems that Big Five and DISC are different approaches that can be used in conjunction (Jones & Hartley, 2013).

Pertaining to Nawaiam, the DISC scores change their names to Assertiveness (Dominance), Sociability (Influence), Tolerance (Steadiness), and Rules (Conscientiousness). Based on the scores and behavioural profile, Nawaiam generates two qualitative reports, one for the company and the other for the applicant, both with tips regarding which kinds of tasks and working conditions are better for the worker. The company has also access to a dashboard that allows matching competence profiles to specific worker characteristics (e.g., if a creative worker is needed, six competence profiles match, but if a creative and adaptable worker is needed, only two out of six profiles match).

Our study wants to advance research on game-based assessments and GRAs in general using Nawaiam to investigate: (1) its capacity to predict four different dimensions of job performance (i.e., task performance, contextual performance, counterproductive work behaviors, and adaptive performance) and its incremental value over traditional measures (i.e., 'Big Five' personality tests); (2) the reactions of applicants to the game, performing a comparison with traditional assessments; and (3) the influence of personal characteristics on game scores and applicant reactions.

Method

Participants

Participants were students and alumni between 18 and 30 years old from a Spanish university. To obtain the study sample, the research team sent an e-mail to the distribution lists of each faculty of the university, informing about the study. Two hundred and ninety-one persons agreed to participate. After providing the information about the study's proposal, 254 were involved. Of them, 244 completed the first part of the study (i.e., the online questionnaire) but only 225 installed and played the GRA (92.2%). Of the 225 players, some of them showed an "inconsistent style" (i.e., the way they play did not allow ascribing them to a behavioral profile, see Variables and Instruments section below). They were invited to play a second time, but 38 either refused or else they again obtained an inconsistent style. Lastly, 5 participants were removed because they were older than 30 and our study is focused on young individuals. Thus, the final sample comprised 182 participants (74.6% of those who initially wanted to participate). Of them, 68.7% were women, 29.7% were men and the rest did not answer or chose "other gender options" (1.6%). The mean age was 21.68 years ($SD = 2.72$). Most participants were undergraduate students ($n = 163$, 89.6%) and used ICT several times a day ($n = 138$, 76.8%). Regarding videogames, 65 (35.7%) did not play videogames, 67 (36.8%) played monthly, and 50 (27.5%) played several times a week or more. Seventy-seven participants (36.8%) had a job, and 23 (12.6%) were actively searching for one. We decided to retain the whole sample in the analyses because Nawaiam was designed to assess people with or without job experience.

Variables and Instruments

Sociodemographic and Work Characteristics

We recorded gender, age, ICT use (5 levels, from *weekly* to *several times a day*), experience with videogames (3 levels, from *none* to *several times a week*), and job experience using an ad hoc survey.

Nawaiam Scores

We used the scores in the dimensions that determine the behavioral profile (assertiveness, sociability, tolerance, and rules) of the participant. Each dimension ranged between 0 and 100. As Nawaiam scores are estimated from the forced choices made in the game, and it is part of the intellectual property of the company that developed the game, we do not have available data to report reliability indices.

Personality

Big Five personality traits were measured using the Spanish version of the BFI-2 (Soto & John, 2017) short form. The instrument uses a five-point rating scale (1 = *strongly disagree* to 5 = *strongly agree*) to measure the following domains: negative emotionality (e.g., "[I am someone who...] is moody, has up and down mood swings"; observed $\alpha = .77$); extraversion (e.g., "is outgoing, sociable"; observed $\alpha = .80$); open-mindedness (e.g., "is curious about many different things"; observed $\alpha = .78$); agreeableness (e.g., "is compassionate, has a soft heart"; observed $\alpha = .63$); and conscientiousness (e.g., "is systematic, likes to keep things in order"; observed $\alpha = .77$).

Performance

Task performance, contextual performance, and counterproductive work behaviors were measured using the Spanish version of the Individual Work Performance Questionnaire (Koopmans, 2015; adapted by Ramos-Villagrasa et al., 2019). This self-report instrument is rated on a Likert scale, ranging from 0 (*seldom*) to 4 (*always*) for task performance (e.g., "I knew how to set the right priorities," 5 items, observed $\alpha = .83$) and contextual performance (e.g., "I took on extra responsibilities," 8 items, observed $\alpha = .83$), and from 0 (*never*) to 4 (*often*) for counterproductive work behaviors (e.g., "I complained about unimportant matters at work," 5 items, observed $\alpha = .68$). Total scores are computed estimating the mean value of each dimension. Adaptive performance was measured with an 8-item scale (Marques-Quinteiro et al., 2015; adapted by Ramos-Villagrasa et al., 2020) ranging from 1 (*totally ineffective*) to 7 (*totally effective*). A sample item is "I find innovative ways to deal with unexpected events". Observed reliability in this study was .83.

Applicants' Reactions

Two types of applicants' reactions were analyzed, all rated from 1 (*strongly disagree*) to 5 (*strongly agree*): perception of Comfort, and perception of Suitability. Both were measured with the Employment Interview Perceptions Scale (EIPS; [Alonso & Moscoso, 2018](#)), adapted for its use with tests and GRAs as an assessment method. The Perception of Comfort subscale includes 5 items (observed $\alpha = .60$ for traditional assessment and $.54$ for game-based assessment). Perception of Suitability has 6 items (observed $\alpha = .63$ for traditional assessment and $.72$ for game-based assessment). Sample items are "I am satisfied with the interview" (comfort) and "The interview will allow me to be evaluated objectively" (suitability). The adapted items can be found in the [Appendix](#).

Procedure

Participants (current and former students) were recruited with the collaboration of the faculties of the researchers' institution. An e-mail with information regarding the study and its purposes was sent, and people interested in participating had to answer the e-mail to be considered. The research team contacted each participant individually by e-mail, informing them about the research procedure, the anonymous treatment of data, and their rights according to the American Psychological Association (APA) standards. If they agreed to participate, a link to a questionnaire for assessing sociodemographic and work characteristics, personality, and applicant reactions to personality measures was sent. After they completed this assessment, they were requested to play the GRA in the following days, completing the questionnaire with applicants' reactions to the game immediately after playing. If their profile was inconclusive, they were asked to play again to receive the report. People who fulfilled all conditions participated in a raffle for one of four vouchers of €50 redeemable at the Amazon website (www.amazon.com).

Analyses

The analyses performed were descriptive statistics (mean, standard deviation), Cronbach's alpha, Spearman's correlations, hierarchical regression analyses, mean comparisons, and analysis of variance (ANOVA). Concerning regression analyses, we used sociodemographic variables, Big Five and GRA scores in Step 3 as predictors. Criteria were job performance. All analyses were performed with Jamovi 2.3.28 (<https://www.jamovi.org/>).

Results

Descriptive statistics are reported in [Table 2](#). The variables displayed the values expected from previous literature (e.g., lower values in Counterproductive behaviors compared with task or contextual performance). Two variables exceeded the acceptable values for kurtosis: job experience (which has a value of 5.310) and Adaptive performance (whose value is 4.820). Thus, in subsequent analyses, non-parametric tests were used.

The associations between variables can be seen in [Table 3](#). Being a woman was associated with less videogame use ($r = -.39, p < .001$). It is remarkable that ICT use was related to one of the reactions to the traditional assessment (i.e., Comfort, $r = .17, p = .020$) but not to any reaction to the GRA. Regarding the GRA scores, only Rules is associated with other dimensions of Nawaiaim, specifically with Assertiveness ($r = -.20, p = .006$) and Tolerance ($r = .22, p = .003$). Talking about the relationship between Nawaiaim and the Big Five, it is noteworthy that only Sociability and Extraversion displayed a significant association ($r = .37, p < .001$). Focusing on its relationship

with the criteria, Task performance and Contextual performance are associated with Sociability ($r = .17, p = .029$ and $r = .19, p = .013$ respectively), and Adaptive performance with Assertiveness ($r = .21, p = .004$), Sociability ($r = .16, p = .028$), and Tolerance ($r = -.17, p = .023$). There is no relationship between counterproductive work behaviors and Nawaiaim scores. On the other hand, Big Five personality traits are mostly related to all dimensions of job performance (mean $r = |.29|$, ranging from $-.39$ to $.54$), with the exception of Open-mindedness with Task performance ($r = .04, p = .585$) and Counterproductive work behaviors ($r = -.06, p = .467$), and Agreeableness with Adaptive performance ($r = .12, p = .104$). To conclude with the correlations, it is worth highlighting the association between scores in Comfort and Suitability with the two assessment methods analyzed (Comfort: $r = .45, p < .001$; Suitability: $r = .50, p < .001$).

Table 2. Descriptive Statistics

Variables	M	SD	Skewness	Kurtosis
Gender	0.71	0.47	-0.745	-1.010
Age	21.68	2.72	1.030	0.705
ICT use	4.75	0.45	-1.370	0.404
Videogame use	0.92	0.79	0.148	-1.390
Job experience	13.50	20.90	2.170	5.310
Assertiveness	39.90	19.00	0.316	0.984
Sociability	68.60	19.90	-0.063	-0.572
Tolerance	66.20	19.00	-0.228	0.576
Rules	60.00	21.10	-0.182	0.220
Negative emotionality	17.30	5.07	-0.068	-0.851
Extraversion	21.30	4.87	-0.362	-0.556
Open-Mindedness	23.10	4.41	-0.545	-0.217
Agreeableness	23.80	3.73	-0.451	-0.301
Conscientiousness	21.20	4.82	-0.157	-0.548
Task performance	2.75	0.82	-0.664	-0.007
Contextual performance	2.31	0.83	-0.115	-0.661
CWB	1.30	0.65	0.212	-0.042
Adaptive performance	44.40	6.51	-1.410	4.820
Comfort – Traditional	18.90	2.53	-0.524	1.070
Comfort – Nawaiaim	19.80	2.36	-0.136	-0.212
Suitability – Traditional	18.00	3.45	-0.130	-0.195
Suitability – Nawaiaim	18.60	3.46	-0.025	0.491

Note. $N = 182$; gender: 0 = men, 1 = women; CWB = counterproductive work behaviors; Traditional = assessment using test/questionnaires.

Concerning regression analyses, results can be seen in [Table 4](#) and [Table 5](#). [Table 4](#) includes the predictive models of performance considering sociodemographic variables (step 1) and game scores (step 2). According with the predictive models, Nawaiaim is involved in Contextual performance (6.6% of explained variance, with Sociability, $\beta = .175, p = .026$, as predictor), and Adaptive performance (11.9% of explained variance, with three Nawaiaim traits as predictors: Assertiveness, $\beta = .244, p = .001$; Sociability, $\beta = .171, p = .023$; and Tolerance, $\beta = -.154, p = .044$).

Continuing with regression analyses, [Table 5](#) includes the predictive models considering Big Five personality traits as step between sociodemographic and game scores. As we can see in [Table 5](#), all models showed the Big Five personality traits to be significant predictors: Task performance (35.6% of explained variance) was determined by Conscientiousness ($\beta = .408, p < .001$) and Negative emotionality ($\beta = -.266, p < .001$); Contextual performance (38.1% of explained variance) was determined by Extraversion ($\beta = .313, p < .001$), Open-mindedness ($\beta = .252, p < .001$), and Conscientiousness ($\beta = .250, p < .001$); Counterproductive work behaviors (14.1%) were explained by Agreeableness ($\beta = -.216, p = .008$) and Negative emotionality ($\beta = .201, p = .029$); and Adaptive performance (12.5% of explained variance) was predicted by Negative emotionality ($\beta = -.178, p = .047$). With the incorporation of GRA scores into the models, only

Table 3. Correlations between Variables

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1. Gender																					
2. Age	.12																				
3. ICT use	.08	.04																			
4. Videogame use	-.39	-.05	-.01																		
5. Job experience	-.02	.40	.02	.11																	
6. Assertiveness	.01	.11	.02	.03	-.01																
7. Sociability	.13	.03	-.06	-.18	.04	-.04															
8. Tolerance	-.02	.06	.07	.09	.01	-.07	-.08														
9. Rules	.11	-.01	.02	.01	.03	-.20	-.06	.22													
10. Negative emotionality	.28	-.06	.02	-.06	-.18	.06	-.01	-.14	-.02												
11. Extraversion	-.05	.04	-.01	-.14	.19	.07	.37	-.09	-.13	-.42											
12. Open-Mindedness	-.13	.06	-.05	.11	.10	.01	.09	-.03	.03	-.02	.16										
13. Agreeableness	.16	.05	.02	-.10	.03	-.13	.05	.09	.04	-.20	.31	.13									
14. Conscientiousness	.21	.01	.02	-.24	.10	-.01	.11	-.09	.10	-.27	.31	.03	.34								
15. Task performance	.14	.13	-.05	-.15	.10	.02	.17	-.05	.03	-.39	.33	.04	.26	.54							
16. Contextual performance	-.05	.10	-.12	-.05	.24	.11	.19	-.06	-.02	-.28	.49	.34	.23	.38	.42						
17. CWB	.01	-.05	.08	.09	-.01	.09	-.10	-.05	-.01	.28	-.19	-.06	-.29	-.19	-.23	-.02					
18. Adaptive performance	.03	.11	.05	-.05	.14	.21	.16	-.17	-.02	-.24	.25	.17	.12	.15	.30	.38	-.02				
19. Comfort – Traditional	.01	.01	.17	-.02	.14	-.07	.13	.12	.01	-.20	.28	.01	.18	.15	.22	.20	-.24	.18			
20. Comfort – Nawaiam	-.07	.01	.06	-.04	.01	-.02	.23	.14	-.04	-.11	.20	.08	.09	.01	.12	.14	-.17	.13	.45		
21. Suitability – Traditional	.04	.03	.12	-.09	.10	.02	.11	.13	.07	-.07	.09	.03	.21	.09	.21	.13	-.19	.05	.40	.15	
22. Suitability – Nawaiam	.04	.11	.06	-.08	.07	-.01	.13	.17	.12	-.09	.06	.12	.22	.10	.14	.11	-.17	.03	.26	.36	.50

Note. N = 182. Bold values are significant associations (up to |.15| = .05; between |.16| and |.23| = .01; higher values = .001); gender: 0 = men; 1 = women; CWB = counterproductive work behaviors.

Table 4. Hierarchical Regression Analyses using Nawaiam as Predictor

	Task Performance			Contextual Performance			Counterproductive Work Behaviors			Adaptive Performance		
	R ²	ΔR ²	p	R ²	ΔR ²	p	R ²	ΔR ²	p	R ²	ΔR ²	p
Step 1	.030			.021			.017			.012		
Step 2	.045	.015	.583	.066	.045	.025	.034	.017	.591	.119	.117	< .001
Coefficients Step 1	β	p		β	p		β	p		β	p	
Gender	.088	.297		-.081	.348		.043	.609		.007	.931	
ICT use	-.039	.613		-.111	.155		.068	.386		.076	.324	
Videogame use	-.111	.187		-.080	.351		.118	.163		-.077	.356	
Coefficients Step 2	β	p		β	p		β	p		β	p	
Gender	.079	.355		-.099	.248		.045	.603		-.021	.793	
ICT use	-.036	.646		-.103	.184		.069	.382		.092	.215	
Videogame use	-.094	.274		-.068	.426		.110	.197		-.060	.456	
Assertiveness	.043	.588		.131	.100		.069	.385		.244	.001	
Sociability	.126	.112		.175	.026		-.074	.353		.171	.023	
Tolerance	-.009	.906		-.024	.766		-.080	.317		-.154	.044	
Rules	-.044	.591		-.012	.884		-.003	.971		.055	.479	

Note. N = 182. Gender: 0 = men, 1 = women; bold values correspond to statistically significant predictors.

one shows incremental variance: Adaptive performance, in which the explained variance increases up to 23.2% (Δ = 10.7%) and has Negative emotionality (β = -.265, p = .003), Assertiveness (β = .255, p < .001) and Tolerance (β = -.213, p = .005) as predictors.

Continuing with the analyses, a comparison of applicants' reactions to traditional assessment and GRA was performed. As can be seen in Table 6, there were significant differences in favor of GRA in Comfort (t = -4.16, p = < .011) and Suitability (t = -2.24, p = .026).

Finally, we used correlations, mean comparisons, and ANOVA to determine possible differences in the game-based assessment scores or applicants' reactions related to personal characteristics: gender, age, studies (undergraduate/graduate), ICT use (up to once a day/more than once a day), videogame use (no/weekly/daily), and working status (not working/working). The analyses only yielded significant differences depending on: (1) working status – Asserti-

veness score of workers was higher than that of non-workers (t = 2.663, p = .008, Cohens' d = -0.387, IC 95% [-0.6880, -0.0854]); (2) ICT use – there are differences in comfort scores regarding traditional assessment in favour of people who use ICT more than once a day (tT = -2.307, p = .022, Cohens' d = -0.360, IC 95% [-0.6610, -0.0559]).

Discussion

Personnel selection is still searching for new assessment methods. In recent years, technology-based methods have become an opportunity. Nevertheless, any new method should maintain or improve the psychometric standards of existing methods (Landers et al., 2021; Salgado et al., 2017; Wiernik et al., 2022). GRAs have the

Table 5. Hierarchical Regression Analyses using the Big Five and Nawaiaam as Predictors

	Task Performance			Contextual Performance			Counterproductive Work Behaviors			Adaptive Performance		
	<i>R</i> ²	ΔR^2	<i>p</i>	<i>R</i> ²	ΔR^2	<i>p</i>	<i>R</i> ²	ΔR^2	<i>p</i>	<i>R</i> ²	ΔR^2	<i>p</i>
Step 1	.030			.021			.017			.012		
Step 2	.356	.326	<.001	.381	.360	<.001	.141	.124	<.001	.125	.113	.001
Step 3	.363	.007	.826	.394	.034	.501	.146	.005	.937	.232	.107	<.001
Coefficients Step 1	β	<i>p</i>		β	<i>p</i>		β	<i>p</i>		β	<i>p</i>	
Gender	.088	.297		-.081	.348		.043	.609		.007	.931	
ICT use	-.039	.613		-.111	.155		.068	.386		.076	.324	
Videogame use	-.111	.187		-.080	.351		.118	.163		-.077	.356	
Coefficients Step 2	β	<i>p</i>		β	<i>p</i>		β	<i>p</i>		β	<i>p</i>	
Gender	.108	.151		-.034	.652		.008	.925		.096	.271	
ICT use	-.041	.518		-.083	.191		.066	.372		.071	.332	
Videogame use	-.011	.881		.019	.785		.082	.314		-.048	.551	
Negative emotionality	-.266	<.001		-.060	.440		.201	.029		-.178	.047	
Extraversion	.047	.539		.313	<.001		-.053	.549		.165	.058	
Open-Mindedness	.033	.619		.252	<.001		-.007	.931		.147	.055	
Agreeableness	.037	.590		.080	.237		-.216	.008		.036	.065	
Conscientiousness	.408	<.001		.250	<.001		-.021	.805		-.013	.876	
Coefficients Step 3	β	<i>p</i>		β	<i>p</i>		β	<i>p</i>		β	<i>p</i>	
Gender	.107	.163		-.035	.646		.011	.900		.083	.318	
ICT use	-.038	.553		-.086	.175		.067	.374		.081	.248	
Videogame use	-.006	.938		.010	.883		.079	.343		-.048	.539	
Negative emotionality	-.286	<.001		-.068	.396		.197	.041		-.265	.003	
Extraversion	.011	.897		.303	<.001		-.044	.650		.079	.384	
Open-Mindedness	.029	.663		.247	<.001		-.006	.936		.133	.069	
Agreeableness	.045	.524		.097	.158		-.207	.013		.093	.230	
Conscientiousness	.406	<.001		.245	.001		-.027	.761		-.060	.472	
Assertiveness	.048	.475		.115	.080		.037	.630		.255	<.001	
Sociability	.063	.374		.008	.909		-.040	.619		.122	.113	
Tolerance	-.019	.775		.004	.951		-.038	.630		-.213	.005	
Rules	-.014	.833		-.007	.921		-.005	.952		.068	.373	

Note. *N* = 182. Gender: 0 = men, 1 = women. Bold values correspond to statistically significant predictors.

potential to do so, but empirical research should prove their value. With the present study, we extend the literature on game-based assessments, a type of GRA that requires more empirical research with respect to its predictive validity, applicant reactions, and the influence of personal characteristics.

Regarding predictive validity, we investigated the relationship between Nawaiaam scores and performance. When the GRA Nawaiaam is the only selection method in predictive models, it can determine contextual performance and adaptive performance. When the Big Five are also considered, some interesting findings should be outlined, supporting the idea that the DISC model and the Big Five should be seen as complementary (Jones & Hartley, 2013). First, Nawaiaam is not involved in the prediction of contextual performance if personality traits are included. Besides that, the explained variance was greater when we used only the Big Five (38.1%) than when we used Nawaiaam (6.6%). Accordingly, we recommend relying solely on personality traits when dealing with contextual performance. With respect to adaptive performance, Nawaiaam demonstrated incremental validity

over the Big Five in predicting adaptive performance, with an approximately 10% increase in the explained variance. This is the first time that any type of GRA provides evidence of predicting this type of performance, which is increasingly in demand in the BANI (brittle, anxious, non-linear, and incomprehensible) work setting. Although the dimensions measured with Nawaiaam are included in some of the other predictive models, there are no significant results that can be generalized to the population. This may be because Nawaiaam is based on the DISC model, whose theoretical and psychometric bases are, to say the least, weak, and previous evidence suggests that the greater the support for the theoretical model behind GRA, the better its results (Ramos-Villagrasa et al., 2022).

Findings about applicants' reactions follow previous literature, obtaining better results for game-based assessments than for traditional assessments. As far as we know, this is the second study to use game-based assessments and the first to measure personality traits. In line with Landers et al. (2021), the more the GRAs look like a game instead of an evaluation, the better it is. It is noteworthy

Table 6. Mean Differences between Traditional Assessment and Serious Game

Applicant reaction	Traditional Assessment <i>M</i>	Game-based Assessment <i>M</i>	<i>t</i>	<i>p</i>	Cohens' <i>d</i>	CI 95%
Comfort	18.90	19.80	-4.16	<.001	-0.1757	-0.3300, -0.0207
Suitability	18.00	18.60	-2.24	.026	-0.3119	-0.4620, -0.1611

Note. *N* = 182.

that a positive effect has been found regarding suitability, as similar research with other GRAs has yielded more favorable results for traditional assessments (Georgiou, 2020). Future research will help to determine whether this is a characteristic of Nawaiam or game-based assessments in general. These results support the idea that using this type of assessment sends signals to candidates that the organization is innovative, which can be a competitive advantage (McChesney et al., 2022). However, given the small effect size found in the present study and the previous study by Landers et al. (2021), it is necessary to consider the contexts in which the use of a game-based assessment can provide an advantage over traditional tests. An example could be in positions where there are very few candidates, such as in the IT sector (Aguado et al., 2019).

Regarding the influence of sociodemographic and work characteristics, we found only a few differences. As for sociodemographic characteristics, our data support the stereotype that women tend to be less involved in videogames; however, this does not have an impact on Nawaiam outcomes or applicant reactions. We found a small effect on applicant reactions, suggesting that people familiar with technology felt more comfortable with the traditional assessments. This may be due to the structure of a traditional assessment applied online, which resembles polls, quizzes, and other common internet activities. No differences were found in terms of age or videogame experience. Thus, our results are in line with previous research claiming that personal characteristics do not substantially impact GRA (Hommel et al., 2022; Melchers & Basch, 2022; Sanchez et al., 2022). Concerning work characteristics, workers tended to score higher than non-workers on assertiveness. This result, which has a small effect size, may reflect the effect of work experience on game decision-making, because workers have had more opportunities to experience circumstances relatively similar to those in Nawaiam. Future research may help to clarify this aspect.

Overall, the present study updates the study of game-based assessment in three ways: (1) it has demonstrated that game-based assessments are capable of predicting adaptive performance; (2) it offers additional evidence regarding positive applicant reactions to game-based assessments, while also noting that the improvement over traditional assessments is marginal; and (3) it highlights that only minor differences exist concerning personal characteristics, which may not have a substantial impact in a real-world context.

Our research contributes to the study of GRAs in general also. The present study is another piece of evidence showing that caution should be taken when using GRAs for personnel selection. As with any assessment method, the specific GRA should demonstrate reliability, construct validity, predictive validity, and be free from bias before its application, regardless of how positive the applicant's reactions are. We must continue researching existing GRAs to draw more conclusions about what works and what does not work in these assessments, and to learn how to build them with improved psychometric characteristics while maintaining what appears to be their greatest strengths (positive reactions and the lack of influence of personal characteristics). The extent to which researchers and professionals can achieve these goals will determine whether the use of GRAs is consolidated as a selection method.

Limitations and Further Research

As any study, the present research has limitations that should be acknowledged. The sample size was small, which limits the generalizability of the results. This is a common limitation in research on GRAs because research on this selection method is still in its early stages (Ramos-Villagrasa et al., 2022). We believe that despite the small sample size, our research was relevant as it delved into one of the less investigated types of GRAs and involved a serious game used in a real setting. Along the same lines, the sample was composed only

of young university students and graduates and not actual applicants. Further research could investigate how this game-based assessment is performed in the personnel selection context.

Another limitation is the use of self-report measures of job performance, because such measures tend to yield more favorable results than other evaluations. However, there are certain scenarios in which self-reports may be useful, such as when it is difficult to collect other data or when the constructs and phenomena under study are still in their infancy (Koopmans et al., 2014). This is the case because some participants are still in the early stages of their careers, and the study of GRAs is just beginning. Further research should extend beyond student samples and self-reports and increase the number of studies involving workers and job applicants with different performance measures.

Considering the aforementioned issues, we must acknowledge threats to the ecological validity as a limitation. There are considerable differences between the context of the evaluation performed in the present study and the actual personnel selection setting. Further research should consider and, if unable to conduct the study in actual selection processes, at least use a simulated context to increase fidelity to reality.

Concerning ideas for future research, more primary studies with different types of GRAs are required (Ohlms et al., 2023). GRA is an umbrella for heterogeneous assessments, so we need to identify which specific game characteristics achieve better reactions, fewer biases, and maintain validity. Meta-analytical studies on GRAs can be conducted when sufficient evidence is available. Additionally, other issues that deserve investigation are more studies about faking of GRAs, the influence of assessment (proctored/unproctored) situation, given that GRAs can be applied remotely, and cross-cultural studies using the same GRAs. All of these will increase our knowledge of when and how to use games in personnel selection processes.

Conflict of Interest

The authors of this article declare no conflict of interest.

References

- Aguado, D., Andrés, J. C., García-Izquierdo, A. L., & Rodríguez, J. (2019). LinkedIn "Big Four": Job performance validation in the ICT sector. *Journal of Work and Organizational Psychology*, 35(2), 53-64. <https://doi.org/10.5093/jwop2019a7>
- Albadán, J., García Gaona, P. A., & Montenegro Marin, C. (2016). Assessment model in a selection process based in gamification. *IEEE Latin America Transactions*, 14(6), 2789-2794. <https://doi.org/10.1109/TLA.2016.7555256>
- Alonso, P., & Moscoso, S. (2018). Employment Interview Perceptions Scale. *Journal of Work and Organizational Psychology*, 34(3), 203-212. <https://doi.org/10.5093/jwop2018a22>
- Andrés, J. C., Aguado, D., & García-Izquierdo, A. L. (2023). Big Four LinkedIn dimensions: Signals of soft skills? *Journal of Work and Organizational Psychology*, 39(2), 75-88. <https://doi.org/10.5093/jwop2023a9>
- Auer, E. M., Mersy, G., Marin, S., Blaik, J., & Landers, R. N. (2022). Using machine learning to model trace behavioral data from a game-based assessment. *International Journal of Selection and Assessment*, 30(1), 82-102. <https://doi.org/10.1111/ijisa.12363>
- 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>
- Gehrig, E., & Bonnstetter, R. (2017). *TTI SI Style Insights vs. Big 5 Personality Inventory: A validation comparison study*. Unpublished paper.
- Georgiou, K. (2021). Can explanations improve applicant reactions towards gamified assessment methods? *International Journal of Selection and Assessment*, 29(2), 253-268. <https://doi.org/10.1111/ijisa.12329>
- Georgiou, K., Gouras, A., & Nikolaou, I. (2019). Gamification in employee selection: The development of a gamified assessment. *International Journal of Selection and Assessment*, 27(2), 91-103. <https://doi.org/10.1111/ijisa.12240>
- 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, Article 106356. <https://doi.org/10.1016/j.chb.2020.106356>
- Gómez, P. L., Barquero, J. D., & Pestana, J. V. (2021). Evidencia empírica del DISC en futuros líderes iberoamericanos: una aportación de William Moulton Marston a la función directiva [Empirical evidence of the DISC in future leaders from Spain and Latin America: A contribution from William Moulton Marston to the managerial role]. *Anuario de Psicología*, 51(3), 130-137. <https://doi.org/10.1344/ANPSIC2021.51/3.30855>
- 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/ijisa.12374>
- Hommel, B. E., Ruppel, R., & Zacher, H. (2022). Assessment of cognitive flexibility in personnel selection: Validity and acceptance of a gamified version of the Wisconsin Card Sorting Test. *International Journal of Selection and Assessment*, 30(1), 126-144. <https://doi.org/10.1111/ijisa.12362>
- Jones, C. S., & Hartley, N. T. (2013). Comparing correlations between four-quadrant and five-factor personality assessments. *American Journal of Business Education*, 6(4), 459-470. <https://doi.org/10.19030/ajbe.v6i4.7945>
- Koopmans, L. (2015). *Individual Work Performance Questionnaire instruction manual*. TNO Innovation for Life – VU University Medical Center.
- Koopmans, L., Bernaards, C. M., Hildebrandt, V. H., de Vet, H. C. W., & van der Beek, A. J. (2014). Measuring individual work performance: Identifying and selecting indicators. *Work*, 48(2), 229-238. <https://doi.org/10.3233/WOR-131659>
- Landers, R. N., Armstrong, M. B., Collmus, A. B., Mujcic, S., & Blaik, J. (2021). 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., Auer, E. M., Collmus, A. B., & Armstrong, M. B. (2018). Gamification science, its history and future: Definitions and a research agenda. *Simulation & Gaming*, 49(3), 315-337. <https://doi.org/10.1177/1046878118774385>
- 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/ijisa.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/ijisa.12376>
- Marques-Quinteiro, P., Ramos-Villagrasa, P. J., Passos, A. M., & Curren, L. (2015). Measuring adaptive performance in individuals and teams. *Team Performance Management*, 21(7-8), 339-360. <https://doi.org/10.1108/TPM-03-2015-0014>
- Marston, W. M. (1928). *Emotions of normal people*. Routledge.
- McChesney, J., Campbell, C., Wang, J., & Foster, L. (2022). What is in a name? Effects of game-framing on perceptions of hiring organizations. *International Journal of Selection and Assessment*, 30(1), 182-192. <https://doi.org/10.1111/ijisa.12370>
- 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/ijisa.12337>
- Nikolaou, I. (2021). What is the role of technology in recruitment and selection? *The Spanish Journal of Psychology*, 24, Article e2. <https://doi.org/10.1017/SJP.2021.6>
- Nikolaou, I., Georgiou, K., & Kotsasarlidou, V. (2019). Exploring the relationship of a gamified assessment with performance. *The Spanish Journal of Psychology*, 22, Article e6. <https://doi.org/10.1017/sjp.2019.5>
- Ohlms, M. L., Melchers, K. G., & Kanning, U. P. (2023). Can we playfully measure cognitive ability? Construct-related validity and applicant reactions. *International Journal of Selection and Assessment*. Advance online publication. <https://doi.org/10.1111/ijisa.12450>
- Owen, J. E., Mahatmya, D., & Carter, R. (2020). Dominance, influence, steadiness, and conscientiousness (DISC) assessment tool. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of personality and individual differences* (pp. 1186-1189). Springer. <https://doi.org/10.1007/978-3-319-24612-3>
- Ramos-Villagrasa, P. J., Barrada, J. R., Fernández-del-Río, E., & Koopmans, L. (2019). Assessing job performance using brief self-report scales: The case of the Individual Work Performance Questionnaire. *Journal of Work and Organizational Psychology*, 35(3), 195-205. <https://doi.org/10.5093/jwop2019a21>
- Ramos-Villagrasa, P. J., Fernández-del-Río, E., & Barrada, J. R. (2020). Does evil prevail? The "bright" and "dark" sides of personality as predictors of adaptive performance. *Sustainability*, 12(2), Article 474. <https://doi.org/10.3390/su12020474>
- Ramos-Villagrasa, P. J., Fernández-del-Río, E., & Castro, Á. (2022). Game-related assessments for personnel selection: A systematic review. *Frontiers in Psychology*, 13, Article 952002. <https://doi.org/10.3389/fpsyg.2022.952002>
- Rodríguez, J. (2004). La tecnología de evaluación ipsativa y el sistema APP-APT de Thomas International [Ipsative Assessment Technology and the Personal Profile Assessment by Thomas International]. *Journal of Work and Organizational Psychology*, 20(2), 169-197.
- Salgado, J. F., Moscoso, S., García-Izquierdo, A. L., & Anderson, N. R. (2017). Inclusive and discrimination-free personnel selection. In A. Arenas, D. Di Marco, L. Munduate, & M. C. Euwema (Eds.), *Shaping inclusive workplaces through social dialogue* (pp. 103-119). Springer International Publishing. https://doi.org/10.1007/978-3-319-66393-7_7
- Sanchez, D. R., Weiner, E., & Van Zelder, A. (2022). Virtual reality assessments (VRAs): Exploring the reliability and validity of evaluations in VR. *International Journal of Selection and Assessment*, 30(1), 103-125. <https://doi.org/10.1111/ijisa.12369>
- Simons, A., Wohlgenannt, I., Zelt, S., Schneider, J., & vom Brocke, J. (2023). Intelligence at play: Game-based assessment using a virtual-reality application. *Virtual Reality*, 27, 1827-1843. <https://doi.org/10.1007/s10055-023-00752-9>
- Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, 113(1), 117-143. <https://doi.org/10.1037/pspp0000096>
- Wiernik, B. M., Raghavan, M., Caretta, T. R., & Coovert, M. D. (2022). Developing and validating a serious game-based assessment for cyber occupations in the US Air Force. *International Journal of Selection and Assessment*, 30(1), 27-47. <https://doi.org/10.1111/ijisa.12378>
- 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>
- Wu, F. Y., Mulfinger, E., Alexander, L., Sinclair, A. L., McCloy, R. A., & Oswald, F. L. (2022). Individual differences at play: An investigation into measuring Big Five personality facets with game-based assessments. *International Journal of Selection and Assessment*, 30(1), 62-81. <https://doi.org/10.1111/ijisa.12360>

Appendix

Adaptation of the applicants' reactions scale by [Alonso & Moscoso \(2018\)](#)

Perception of Comfort

Estoy satisfecho con mis resultados en la prueba/el juego.

He estado motivado/a durante la prueba/el juego para conseguir el mejor resultado posible.

Me ha parecido difícil contestar a las preguntas/jugar al juego.*

Me he sentido cómodo/a con las preguntas de la prueba/jugando al juego.

La prueba/El juego respeta mi privacidad.

Perception of Suitability

La prueba/El juego me permite falsear mis respuestas.*

La prueba/El juego me parece justa/o.

La prueba/El juego permite que se me evalúe de forma objetiva.

Los candidatos mejor evaluados con la prueba/el juego tendrán mejor desempeño.

La prueba/El juego es adecuada/o para decidir cuál es el mejor candidato.

Esta prueba/Este juego facilita la toma de decisiones a los seleccionadores.

Note. Items labeled with * are reversed.