

# THE SCIENCE BEHIND OWIWI

*AN EXPLANATORY MANUAL ON OUR  
SCIENTIFIC PROCESS*



# DISCLAIMER

---

The game presented by Owiwi has been developed following the most recent rigorous methodology in psychological construct development. The scientific team of Owiwi followed the methodology of Situational Judgement Tests (SJTs) in developing the assessment behind Owiwi.

**THE PURPOSE OF THIS MANUAL IS TO PROVIDE A CLEAR PICTURE OF ALL THE PROCESSES INVOLVED IN THE CONSTRUCTION OF A NEW PSYCHOMETRIC TOOL THAT MEASURES SOFT SKILLS. ALL THE ETHICAL GUIDELINES WERE FOLLOWED WITH THE AIM TO OFFER A SCIENTIFICALLY WELL-ESTABLISHED MEASURE.**

The Owiwi Soft Skill Manual is proprietary to Owiwi Private Company and no ownership rights are hereby transferred. No part of the Manual shall be used, reproduced, translated, converted, adapted, stored in a retrieval system, communicated or transmitted by any means, for any commercial purpose, including without limitation, sale, resale, license, rental or lease, without the prior express written consent of Owiwi P.C.



**Dr. Nikolaou is a Work & Organizational Psychologist, Associate Professor in Organisational Behaviour and Director of the MSc in Human Resources Management at Athens University of Economics and Business.**

**He has written the books “Organizational Psychology & Behavior” (with Maria Vakola) and “Managing Human Capital – Greek Case Studies” and co-edited with Janneke Oostrom the book Employee Recruitment, Selection, and Assessment. Contemporary Issues for Theory and Practice. (Routledge/Psychology Press). He has also published extensively in peer-reviewed scientific journals while his research interests lay in the field of Organisational Behaviour and Human Resources Management, and more specifically in employee recruitment, selection and assessment; while maintaining active links with the industry through consulting projects and executive training.**

**He is a member of the Academy of Management, Society for Industrial and Organizational Psychology, European Association of Work and Organizational Psychology and he is also the co-founder of the European Network of Selection Researchers (ENESER).**

**Taking into account his expansive expertise, industry knowledge and skill set, Dr. Nikolaou is Owiwi’s Chief Science Officer and we consider him a vital and integral component of our endeavors. He is responsible for all scientific matters pertaining to the development of our tool as well as for providing key insights and trends in the HR sector.**

# WHAT IS A SITUATIONAL JUDGEMENT TEST (SJT)?

---

SJTs are a popular personnel selection method, designed to assess an applicant's judgment regarding a situation encountered in the workplace (Weekley & Ployhart, 2006). Their popularity is based on the assertion that they assess soft skills and job-related skills not tapped by other measures, with a low adverse impact that nurtures positive applicant reactions. SJTs present respondents with work-related situations and a list of plausible courses of action. Respondents are asked to evaluate each course of action for either the likelihood that they would perform the action or the effectiveness of the action (Whetzel & McDaniel, 2009). Thus, SJTs tend to determine behavioral tendencies, assessing how an individual will behave in a certain situation, and knowledge instruction, which evaluates the effectiveness of possible responses.

# THE PREDICTIVE VALIDITY OF SJTS

---

Several studies (e.g., McDaniel & Nguyen, 2001) have demonstrated the predictive validity of SJTs. McDaniel, Hartman, & Whetzel, & Grubb (2007) demonstrated in their meta-analysis that SJT scores have an average observed validity of .20, and have incremental validity over cognitive ability scores and Big Five personality ratings. Christian et al. (2010) meta-analytically showed that video-based SJTs have higher validity than paper-and-pencil SJTs for predicting interpersonal skills. That is, video-based SJT scores of interpersonal skills had an average validity of .47, which was significantly higher than the average validity of .27 for paper-and-pencil SJT scores of interpersonal skills. Video technology has been successfully applied to SJTs (e.g., Olson-Buchanan & Drasgow, 2006).

**ONE ADVANTAGE OF VIDEO-BASED SJTS IS THAT THE INCREASED FIDELITY OF PRESENTING THE SITUATIONS IN VIDEO FORMAT MIGHT LEAD TO HIGHER PREDICTIVE VALIDITY WHEREAS SJTS' HIGHER REALISM MIGHT RESULT IN MORE FAVORABLE APPLICANT REACTIONS (LIEVENS & SACKETT, 2006).**

# SJT DEVELOPMENT

---

There are two popular methods for developing SJT items: critical incident and theory-based methods (Weekley, Ployhart, & Holtz, 2006). The scientific team of Owiwi followed the critical incident method in developing the SJTs item stems and response options.

The critical incident method (Flanagan, 1954) is the most common approach used to identify the content of the items (Motowidlo, Hanson, & Crafts, 1997). The critical incidents can be collected from archival records or from interviews with subject matter experts (SMEs), for example managers, incumbents, clients, or other key stakeholders. The antecedents, or situational descriptors of the context leading up to the incident, are used to develop the item stem (scenario) while the subsequent behavior described is used in the development of one or more of the response options.

The development process we adopted followed three stages (Motowidlo et al., 1990).

**STAGE 1: Competencies identification**

---

**STAGE 2: Development of SJT stems (scenario)**

---

**STAGE 3: Development of SJT stems' response options**

---

---

## STAGE 1: Competencies identification

Owiwi's scientific team carried out extensive literature research in order to identify the core competencies / skills that organizations are seeking from young recruits. We paid special attention to the necessary competencies organizations look for among university graduates especially, e.g. in graduate recruitment. A list of the core competencies / skills was created and four of them were chosen for the first version of the SJT and the game which are the competencies of

- ▲. resilience,
- ▲. adaptability,
- ▲. **willingness to change,**
- ▲. decision-making.

Definitions of resilience, adaptability, **willingness to change** and decision-making originated from theory and previous empirical research in the fields of management and psychology.

- ▶ In non-organizational contexts, resilience is defined as “a class of phenomena characterized by patterns of positive adaptation in the context of significant adversity or risk” (Masten & Reed, 2002, p. 75). In an organisational context, resilience is defined as “the developable capacity to rebound or bounce back from adversity, conflict, and failure or even positive events, progress, and increased responsibility” (Luthans, 2002, p. 702). In other words, resilience refers to a dynamic process encompassing positive adaptation within the context of significant adversity.
- ▶ The majority of researchers agree that “adaptability is related to change and how people deal with it; that is to say, people’s adjustment to changing environments” (Hamtiaux, Houssemand, & Vrignaud, 2013, p. 130).

- ▶ **Willingness to change is defined as “a positive behavioural intention towards the implementation of modifications in an organization’s structure, or work and administrative processes, resulting in efforts from the organization member’s side to support or enhance the change process” (Metselaar, 1997, p. 42).**
- ▶ **Decision-making is defined as an intellectual process leading to a response to circumstances through selection among alternatives (Nelson 1984). Competent decision making requires several key skills including the ability to understand information, integrate information in an internally consistent manner, identify the relevance of information in a decision process, and inhibit impulsive responding. Performance on these skills is expected to reflect the degree of congruence between characteristics of the decision maker and the demands of the task and context (Finucane et al., 2005).**

The next four competencies / skills that were chosen for the second version of the SJT and the game are the competencies of

- ▲. teamwork,
- ▲. learning agility,
- ▲. accountability,
- ▲. integrity.

▶ Teamwork refers to the extent to which a team member is able to meet the team’s output goals (e.g., quality, functionality, and reliability of outputs), the expectations of other members, or its cost and time objectives (Ancona & Caldwell, 1992).

▶ Learning agility is defined as “the willingness and ability to learn new competencies in order to perform under first-time, tough, or different conditions” (Lombardo & Eichinger, 2000, p. 323).

▶ Accountability is the fundamental social contingency driving one’s behavior and decisions. It is defined as a “perceived expectation that one’s decisions or actions will be evaluated by a salient audience and that rewards or sanctions are believed to be contingent on this expected evaluation” (Hall & Ferris, 2011, p. 134).

▶ Integrity reflects a proclivity to engage in just/moral behavior. It is described by a set of beliefs, attitudes and actions reflecting one’s personal values and mores (Meriac, Woehr, & Banister, 2010).

## STAGE 2: Development of SJT stems (scenario)

In the second stage of the SJT development Owiwi adopted the critical incident technique with subject matter experts - experienced employees in various organizations (HR Directors & HR Managers, and Recruiters). Thirty semi structured interviews were conducted to identify critical incidents in each domain definition that we identified. More specifically, participants were asked to recall exceptionally good and exceptionally poor examples of performance for each one of the eight core competencies/skills we described to them. Subsequently, experienced researchers and academics selected the best non-redundant critical incidents from the total pool and rewrote them into stems-scenarios of similar length and format, which hereinafter were called SJT scenarios. The result of this stage was 157 real case scenarios, which should be subjected to face and content validity procedures (Chan & Schmitt, 1997).

Two researchers independently produced a set of possible answers to each scenario, following the guidelines of the subject matter experts of the previous stage, including: two neutral response items, one that predicts performance (positive response) and one response item that diverts from performance (negative response).

## STAGE 3: Development of SJT stems' response options

Subsequently, Owiwi needed to determine the response instructions to the SJT items. There are two types of SJT response instructions: Knowledge-based response instructions, also known as 'should-do' response instructions, ask the test taker to identify the best or correct course of action in the given situation. Behavioral tendency response instructions, also known as 'would-do' response instructions, ask the test taker to express how he or she would likely behave in the given situation (McDaniel, et al. 2007).

The two instruction types relate to the distinction between typical and maximal performance (Cronbach, 1984). Maximal performance tests assess test takers' performance when doing their best and are generally used to make inferences about ability. Typical performance tests assess how test takers typically behave and are generally used to make inferences about personality, attitudes, and other non-cognitive aspects. SJTs with knowledge response instructions are maximal performance tests as test takers make judgments about what constitutes effective performance. SJTs with behavioral tendency response instructions are typical performance tests as test takers report how they typically behave (McDaniel et al., 2007).

▶ **“WOULD DO” INSTRUCTIONS WERE CHOSEN, SINCE THE SJT ASSESSES INTERPERSONAL SKILLS AND NOT ABILITIES AND WE ARE ALSO MOSTLY INTERESTED IN APPLICANTS’ TYPICAL BEHAVIOR, SINCE THE GAME WILL BE USED IN RECRUITMENT/SELECTION SETTINGS AND TRAINING AND DEVELOPMENT. SUBSEQUENTLY, ALL SCENARIOS AND RESPONSES ITEMS WERE PRESENTED TO A FIRST GROUP OF SUBJECT MATTER EXPERTS WHICH WERE ASKED TO EVALUATE AND COMMENT ON THE SCENARIOS, INDICATING THE MOST POSITIVE AND THE MOST NEGATIVE RESPONSE ITEM IN ORDER TO FORMULATE THE FINAL SET OF SCENARIOS UNDER CONTENT VALIDATION. AS A RESULT, 18 ITEMS - SCENARIOS ON AVERAGE FOR EACH CONSTRUCT WERE PRODUCED WITH FOUR RESPONSE OPTIONS FOR EACH SCENARIO.**

# SJT VALIDITY

---

A critical aspect of any measure, especially new, is to establish its validity. Validity is about responding to questions, such as the following:

- ▲. Is the measure accurate and relevant? (content validity)  
Does it assess what it is supposed or it says that it assesses
- ▲. (construct validity)  
Does it predict work outcomes, such as job performance?
- ▲. (predictive or criterion-related validity)

## Content Validity

The first step in establishing the validity of the SJT is to explore its content validity. Content validation simply asks the question, “is the content of the test relevant to the characteristic being measured? (Hammond, 1995). Content validation procedures are important when developing a measure since it is necessary to construct items that sample the domain in question. During this stage, the questionnaire has been administered to eight subject matter experts, who were asked to sort all four responses from the best alternative to the worst one. They were also asked to rate their effectiveness, as well. Cohen’s Kappa was used to check if the identification of best answers among them was acceptable. As follows, their consensus was used to proceed to the extraction of the final scenarios to be tested for construct validity along with the most appropriate scoring key.

**AFTER SEVERAL VALIDATION AND FACTOR ANALYTIC PROCEDURES THAT ENABLED US TO ASSUME THE OPTIMUM FIT TO DATA (KLINE, 2005), WE REACHED A FINAL POOL OF 32 SCENARIOS IN TOTAL (4 PER CONSTRUCT).**

## Construct validity

The next important step in the development of the SJT is to establish its construct validity (Nunnally & Bernstein, 1994). Construct validity responds to the question, “does the measure actually measure what it claims to measure?” This is a very important issue for every new measure/assessment. In order to explore the construct validity of the SJT, the Owiwi research team conducted a validation survey. **A total of N=938 participated in this survey, mostly recent graduates and/or newcomers in organizations.** They were asked to complete an on-line version of the SJT along with a number of well-established measures of the four skills in question.

## More specifically, we used the following measures:

- 1** The Resilience Scale by Wagnild & Young (1993) has been used to explore the construct validity of the SJT resilience scenarios. The Resilience Scale contains 25 items using a 7-point scale from 1 (Strongly Disagree) to 7 (Strongly Agree). Its Alpha reliability in our study was .89. The minimum acceptable level of Alpha reliability according to Nunnally & Bernstein (1994) is .60;
- 2** To measure adaptability, the Adaptability Scale of Martin, Nejad, Colmar, & Liem (2012) was used. The Scale consists of 9 items (e.g., “I am able to think through a number of possible options to assist me in a new situation”) using an 1 (‘strongly disagree’) to 7 (‘strongly agree’) scale. Its Alpha reliability in our study was .89
- 3** **Willingness to change** was measured using items from the HEXACO Personality Inventory (Lee & Ashton, 2004). The scale contains 10 items (e.g., “I react strongly to criticism”) using a 5-point scale, from 1 (‘strongly disagree’) to 5 (‘strongly agree’). Its Alpha reliability in our study was .74
- 4** Decision making was measured with an adopted version of the Mincemoyer and Perkin (2003) employer’s decision-making scale which assesses factors, such as defining the problem; generating alternatives; checking risks and consequences of choices; selecting an alternative; and evaluating the decision. Each factor consisted of three to five items (e.g., “I easily identify my problem”). Its Alpha reliability in our study was .77.

**5** Teamwork was measured using the Teamwork Quality (TWQ) scale (Meslec, Nicoleta, Curseu, Petru, Lucian, 2015). The TWQ scale contains 10 items (e.g., “The group members feel they belong to this group” and a 5-point scale, from 1 (‘strongly disagree’) to 5 (‘strongly agree’). Its Alpha reliability in our study was .94.

**6** Learning agility was measured using items from the HEXACO Personality Inventory (Lee & Ashton, 2004). The love of learning scale contains 10 items (e.g., “I am thrilled when I learn something new”) using a 5-point scale, from 1 (‘strongly disagree’) to 5 (‘strongly agree’) scale. Its Alpha reliability in our study was .79.

**7** Accountability was measured using Mero, Neal, Guidice, Rebecca, and Werner’s (2014) perceived accountability scale. The perceived accountability scale contains 6 items (e.g., “I am required to justify or explain my performance in terms of achieving unit goals”) using a 5-point scale, from 1 (‘strongly disagree’) to 5 (‘strongly agree’). Its Alpha reliability in our study was .87.

**8** Integrity was measured using the subscale “morality/ethics” of the multidimensional work ethic profile (Meriac, Woehr, Gorman, Thomas, 2013). The morality/ethics subscale contains 4 items (e.g., “People should be fair in their dealings with others”) using a 5-point scale, from 1 (‘strongly disagree’) to 5 (‘strongly agree’). Its Alpha reliability in our study was .88.

The subsequent statistical analysis confirmed the structural validity of the SJT (Nunnally & Bernstein, 1994; Robinson et al., 1991), along with its convergent and divergent validity (Campbell & Fiske, 1959), confirming the psychometric qualities of the SJT.

More specifically after Confirmatory Factor Analytic methodology is applied, all skills seem to correlate with each other at a low to moderate level, but still statistically significant (from .290 to .558), that is a sign of convergence. Signs of discriminant validity have been appeared also after Factor Analytic techniques had been applied, assuming that all skills are not sharing statistically significant variance with other well established measurements (correlations ranging from -.101 to .307, all statistically significant at  $p < .05$ ) designated to measure completely irrelevant skills (e.g.willingness to change vs team work). At all cases the model fit to data is acceptable by current literature (Chen, Curran, Bollen, Kirby, & Paxton, 2008).

**IN OTHER WORDS, THE ADDITION OF GAME ELEMENTS TO THE SJT AND ITS CONVERSION INTO AN ADVENTURE ONLINE STORY CONFIRMS THE CONSTRUCT VALIDITY OF THE CURRENT GAMIFIED METHOD, THAT IS AMONG THE FIRST VALIDATED INSTRUMENTS USING GAME ELEMENTS IN ORDER TO ASSESS CANDIDATES' SOFT SKILLS (GEORGIU, GOURAS, NIKOLAOU, 2019).**

## Norming

The norming procedure is based on the guidelines of stratified sampling (Angoff, 1984), which “breaks up” the sample to layered sub samples based on particular “strata”, i.e. demographic data gathered throughout the game. There is a variety of criteria used to set up a specialised norm depending mainly on our norming strategy.

Literature and practice led us to accept the norming standards of a sub sample no less than 300 incumbents. Having that in mind, we statistically analysed the whole sample in terms of declared level (junior, intermediate, senior), age groups, education level, status (employee or not), gender, sector and country and we generate statistical distributions for each norm.

The norming sample is comprised of 5371 participants all over the world. These participants completed the short version of the game of 4 islands (resilience, willingness to change, adaptability and decision making) during 2017 till now. The norming procedure for the next 4 islands (i.e. skills) is in progress. The segmentation of the sample is as such (only the top choices): IT: 10,7%, engineering: 7,7%, Financial service: 6,2%, Energy: 5,8%, Other: 27% etc. Almost 50% of the sample is Turkish and 35% is Greece. 38% of the sample are graduates and 36,5% are master holder. There is also a 15,5% that are PhDs. Approximately 35% of them are students and 28% are employees. 25% of the sample is 25 years old and younger while 50% is more than 25 years old and less than 30 years old. Example: we would like to benchmark all BSc holders who are students yet. This leads us to 1025 incumbents and set of statistics (e.g. Q1, D1, Median, Q3, D9) for each game dimension and overall score.

Recently, Owiwi has digitalized the norming procedure so as to cover customised clients’ needs and identify the proper norm each time an individual report is generated taking advantage of the benefits of artificial intelligence.

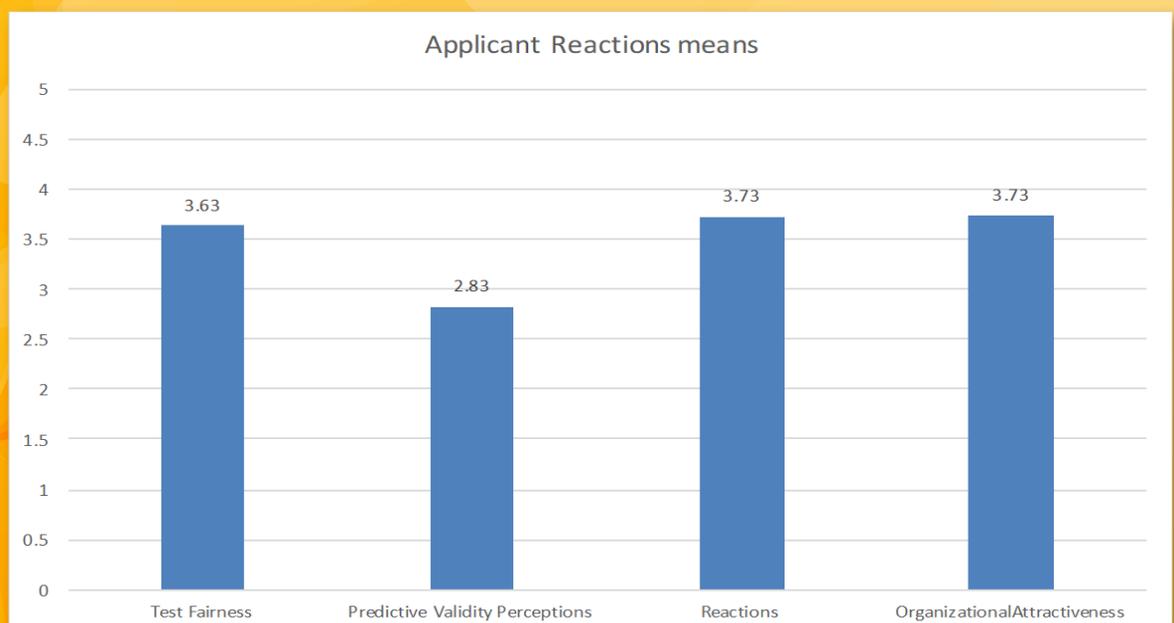
## Predictive Validity

There is preliminary evidence supporting the predictive validity of OWIWI's game. More specifically, the gamified assessment method is found to be associated with self-reported measures of performance, such as job performance and GPA (Nikolaou, Georgiou, Kotsasarlidou, 2019). Moreover, it can predict academic performance (i.e., GPA), above and beyond traditional selection methods, such as cognitive ability and personality tests (Nikolaou, Georgiou, Kotsasarlidou, 2019).

## Applicant Reactions – 1<sup>st</sup> Survey

Having provided preliminary evidence to support the construct and predictive validity of the game, OWIWI had to explore whether the game yields positive reactions to applicants as well.

The findings of our 1st applicant reactions survey support that OWIWI’s game creates positive experiences among users. More specifically, using a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree), N=131 players (mean=26yrs, sd=5.3, 65% females, 71% graduates/postgraduates) described the game as fair and valid and indicated positive applicant reactions, considering for example that it is appropriate for an employer to administer such a game and that a company using this game would be a good place to work (Nikolaou & Georgiou, 2017).



## Applicant Reactions – 2<sup>nd</sup> Survey

To support further the relationship between OWIWI's game and positive applicant reactions, we examined not only the effect of our gamified assessment method on applicant reactions and outcomes, but also this effect versus more traditional methods.

Overall, the findings of our 2nd applicant reactions survey demonstrate that OWIWI's game is more attractive to candidates than traditional selection methods.

The current gamified assessment method might increase applicants' process satisfaction and their attraction to organization to a greater extent than traditional selection methods. Moreover, applicants might consider an organization using the gamified assessment method more attractive as employer, which in turn, leads to better recruitment outcomes. More specifically, OWIWI's game is found to have a stronger effect on applicants' organizational attractiveness and in turn, recommendation intentions than a traditional SJT, when applicants have high levels of video gaming experience (Gkorezis, Georgiou, Nikolaou, & Kyriazati, under review).

# SUMMING UP

---

The aforementioned analysis demonstrates the development steps the Owiwi team undertook in order to establish the validity of the SJT, which forms the basis of the game. These were necessary steps and one could say the minimum steps in order to launch the game. However, Owiwi's team has provided preliminary evidence to the following important issues as well:

- 1 The equivalence of the game with the SJT
- 2 The stability-reliability of the game and how applicants perceive it
- 3 The predictive validity of the game

# REFERENCES

---

- Ancona, D. G., Caldwell, D. F. 1992. Demography and design: Predictors of new product team performance. *Organization Science*, 3: 321-341.
- Bhattacharya, M., Gibson, D. E., & Doty, D. H. (2005). The effects of flexibility in employee skills, employee behaviors, and human resource practices on firm performance. *Journal of management*, 31(4), 622-640.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and Discriminant Validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment is situational judgment tests: subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, 82(1), 143-159.
- 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.
- Cronbach, L. J. (1984). *Essentials of psychological testing* (4th ed.). New York, NY: Harper & Row.
- Finucane ML, Mertz CK, Slovic P, Schmidt ES. Task complexity and older adults' decision-making competence. *Psychology and Aging*. 2005;20(1):71-84.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological bulletin*, 51(4), 327-358
- Georgiou, K., Gouras, A., Nikolaou, I. (2019). Gamification in employee selection: The development of a game-based assessment. *International Journal of Selection and Assessment*, 27, 91-103.

Golden, W., & Powell, P. (2000). Towards a definition of flexibility: in search of the Holy Grail? *Omega*, 28(4), 373-384.

Gkorezis, P., Georgiou, K., Nikolaou, I., Kyriazati, A. (under review). Gamified or situational judgement test? A moderated mediation model of recommendation intentions. *European Journal of Work and Organizational Psychology*.

Hall, A. T., & Ferris, G. R. (2011). Accountability and extra-role behavior. *Employee Responsibilities and Rights Journal*, 23, 131-144.

Hammond, S. (2000). Using psychometric tests. In G. Breakwell, S. Hammond and C. Fife-Schaw (Eds.) *Research methods in psychology*, p.p. 175-193, London: Sage.

Hamtiaux, A., Houssemand, C., & Vrignaud, P. (2013). Individual and career adaptability: Comparing models and measures. *Journal of Vocational Behavior*, 83(2), 130-141.

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

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, 39(2), 329-358.

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

Lievens, F., & Coetsier, P. (2002). Situational Tests in Student Selection: An Examination of Predictive Validity, Adverse Impact, and Construct Validity. *International Journal of Selection and Assessment*, 10(4), 245-257.

Lombardo, M. M., & Eichinger, R. W. (2000). High potentials as high learners. *Human Resource Management*, 39, 321-330.

Luthans, F. (2002). The need for and meaning of positive organizational behavior. *Journal of Organizational Behavior*, 23(6), 695-706.

Martin, A. J., Nejad, H., Colmar, S., Liem, G. A. D., & Collie, R. J. (2015). The role of adaptability in promoting control and reducing failure dynamics: A mediation model. *Learning and Individual Differences*, 38, 36-43.

Masten, A., & Reed, M. (2002). Resilience in development. [In:] CR Snyder, SJ López (Eds.), *Handbook of positive psychology* (pp. 74-88): New York: Oxford University Press. McCubbin, HI, Thompson, EA, Thompson, AI, & Futrell, J.(1999). *The dynamics of resilient families*. London: SAGE Publications.

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. L., III. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, 60, 63-91.

McDaniel MA, Nguyen NT. (2001). Situational judgment tests: a review of practice and constructs assessed. *International Journal of Selection and Assessment*.9:103-13.

Meriac, J. P., Woehr, D. J., Gorman, C. A., Thomas, A. L. E. (2013). Development and Validation of a Short Form for the Multidimensional Work Ethic Profile. *Journal of Vocational Behavior*, 82(3), 155-164.

Mero, Neal, P., Guidice, Rebecca, M., & Werner, Steve (2014). A field study of antecedents and performance consequences of perceived accountability. *Journal of Management*, 40(6), 1627- 1652.

Meslec, Nicoleta, Curseu, Petru, Lucian (2015). Are balanced groups better? Belbin roles in collaborative learning groups. *Learning and Individual Differences*, 39, 81-88.

Metselaar, E.E. (1997). *Assessing the Willingness to Change: Construction and Validation of the DINAMO'*, PhD Thesis. Amsterdam: Free University of Amsterdam.

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

Motowidlo, S.J., Hanson, M.A. and Crafts, J.L. (1997) Low-fidelity simulations. In D.L. Whetzel and G.R. Wheaton (eds.), *Applied Measurement Methods in Industrial Psychology*. Palo Alto, CA: Davies-Black.

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, USA.

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

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

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

Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (1991). Criteria for scale selection and evaluation. In J. P. Robinson, P. R. Shaver & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes*. San Diego: Academic Press.

Wagnild, G. M., & Young, H. (1993). Development and Psychometric Evaluation of the Resilience Scale. *Journal of nursing measurement*, 1(2).

Weekley, J. A., & Ployhart, R. E. (2006). An introduction to situational judgment testing. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests*, Mahwah, NJ: Lawrence Erlbaum Associates.

Weekley, J. A., Ployhart, R. E., & Holtz, B. C. (2006). On the development of situational judgment tests: Issues in item development, scaling, and scoring. *Situational judgment tests: Theory, measurement, and application*, 26, 157-182.

Whetzel, D. L., & McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review*, 19(3), 188-202.

**THANK YOU!**

We hope you enjoyed the reading!

**Do you have any questions? Let's talk!**

**Send us an email to: [hello@owivi.co.uk](mailto:hello@owivi.co.uk)**

**Here are also our recent whitepapers.**



**Use the link below to read it**  
**<https://bit.ly/2ZKtNAD>**



**Use the link below to read it**  
**<https://bit.ly/2GRQmfd>**