owiwi

THE SCIENCE BEHIND OWIWI

Introduction and 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 Judgment 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.

Dr. Nikolaou

Dr. Nikolaou is a Work & Organizational Psychologist and an Associate Professor in Organizational Behavior at Athens University of Economics and Business (AUEB), Department of Management Science and Technology.

Dr Nikolaou has carried out his postgraduate studies (MSc, PhD) as a scholar of State Scholarships Foundation of Greece (IKY) at Manchester School of Management, University of Manchester Institute of Science and Technology (UMIST). He has gained wide working experience as an Assistant Manager for PricewaterhouseCoopers, Greece at the department of Global Human Resources Solutions and as Head of the Training Department at Egnatia Bank before starting his academic career.

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).

Dr. Nikolaou is Owiwi's Chief Science Officer, 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. With his extensive expertise, industry knowledge and skill set, Dr. Nikolaou is a vital and integral component of our endeavors.

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. SJT's 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.

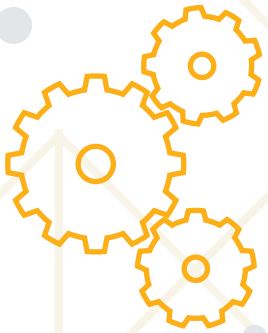
The Predictive Validity of SJTs

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.

Our development process
see Motovidio et el., 1990

STAGE 1: Competencies identification

STAGE 2: Development of SJT stems (scenario)

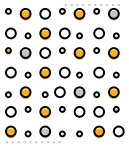
STAGE 3: Development of SJT stems' response options

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

Flexibility

Decision-making

STAGE 1: Competencies identification

Definitions of resilience, adaptability, flexibility 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 organizational 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).

- Flexibility is defined as the capacity to adapt (Golden & Powell, 2000). *"Employee behavior flexibility represents adaptable as opposed to routine behaviors; it is the extent to which employees possess a broad repertoire of behavioral scripts that can be adapted to situation-specific demands"* (Bhattacharya, Gibson, & Doty, 2005, p.624).

- 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).

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). Twenty 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 four 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 104 real case scenarios (26 scenarios per construct on average), 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, 23 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)**

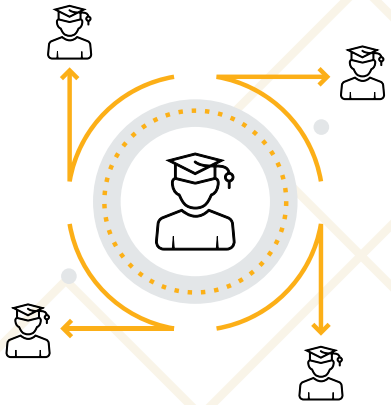
Content Validity

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. The final pool of items that survived this process was 32 scenarios in total (8 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. Participants (N=321) in this survey were (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 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;

3 Flexibility was measured using items from the HEXACO Personality Inventory (Lee & Ashton, 2004). The Flexibility scale contains 10 items (e.g., “I react strongly to criticism”) using a 5-point scale, from 1 (‘strongly disagree’) to 5 (‘strongly agree’) scale. Its Alpha reliability in our study was .74

4 Decision making was measured with an adopted version of the Mincemoyer and Perkin (2003) employed’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.

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.

Summing up and next steps

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, it is crucial to move on quickly with the next steps, as well. The studies we have designed and executed will subsequently try to establish the following important issues:

1. The equivalence of the game with the SJT (1st study)
2. The stability-reliability of the game and how applicants perceive it (2nd study)
3. The predictive validity of the game (3rd study)

References

Bhattacharya, M., Gibson, D. E., & Doty, D. H. (2005). The effects of flexibility in employee skills, employee behaviours, 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.

Christian, M. S., Edwards, J. C., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. Personnel Psychology, 63, 83-117.

Cronbach, L. J. (1984). Essentials of psychological testing (4th ed.). New York, NY: Harper & Row.

Finucane ML, Mertz CK, Slovic P, Schmidt ES. (2005). Task complexity and older adults' decision-making competence. Psychology and Aging, 20, 71–84.

Flanagan, J. C. (1954). The critical incident technique. Psychological bulletin, 51, 327-358

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

Hammond, S. (2000). Using psychometric tests. In G. Breakwell, S. Hammond and C. Fife-Schaw (Eds.) Research methods in psychology (pp. 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.

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. Multivariate Behavioral Research, 39, 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, 1181-1188

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, 245–257.

Luthans, F. (2002). The need for and meaning of positive organizational behavior. Journal of Organizational Behavior, 23, 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. [8]

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.

Mincemoyer, C. C., & Perkins, D. F. (2003). Assessing decision-making skills of youth. Paper presented at 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.
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. 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), 165-178.

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. In J. A. Weekley & R. E. Ployhart (Eds.), Situational judgment tests, Mahwah, NJ: Lawrence Erlbaum Associates.

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