



## **Indirect Skill Assessment Using AI Technology**

**Maria Tountopoulou**

Ison Psychometrica, Athens, Greece

**Fotini Vlachaki**

Ison Psychometrica, Athens, Greece

**Petros Daras**

Visual Computing Lab Information Technologies  
Institute Center for Research & Technology Hellas, Thessaloniki, Greece

**Nicholas Vretos**

Visual Computing Lab Information Technologies  
Institute Center for Research & Technology Hellas, Thessaloniki, Greece

**Aristeidis Christoforidis**

Visual Computing Lab Information Technologies  
Institute Center for Research & Technology Hellas, Thessaloniki, Greece

### **ABSTRACT**

**In the new context of transitional, mobilised and globalised labour markets, an urgent need has emerged for meaningful assessment tools, methods and techniques to measure and recognize the workers' skills. This paper aims to present a novel approach in skill assessment developed under NADINE H2020 project<sup>1</sup>, the indirect skill assessment, as an alternative or supplement to the traditional selection methods. AI technology is utilised to facilitate the indirect skill assessment via two agnostic content serious games (Tetris and 2048), based on the evaluation of an individual's performance in playing a game outside a situational judgement test (SJT) context. Novel datasets have been developed, comprised of game sessions and the corresponding skill assessments of the players through validated psychometric questionnaires, which were the basis for the algorithm training that would provide the estimation of a player's skillset. The trained neural models for both games proved to have strong skill assessment capabilities, indicating that there is indeed a correlation between a person's action sequence and his/her different skills.**

**Keywords:** Indirect skill assessment; AI technology in assessment, agnostic content games; serious games

---

<sup>1</sup> This article is part of the NADINE project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 822601.

## INTRODUCTION

Skills are recognised as extremely important in the modern economy and in everyday life, especially within the on-going structural changes in the global labour market and rapid technological change, as well as demographic changes, with ageing populations, and workforce's migration. At the same time, the supply of skills often struggles to match these rapidly evolving demands. In this respect, special emphasis is placed at European and international level on skills related to adult labour market participation, employability skills and life skills, while there are different perceptions and approaches to the definition and assessment of 'skills'.

The new context of transitional, mobilised and globalised labour markets stresses an urgent need for meaningful relevant assessment tools, methods and techniques to measure and recognize the workers' skills. The skills-based profiling tools are generally considered to improve the matching process, in terms of quality, efficiency, and potential for automation, enabling employers to acquire more detailed information on competences, technical, generic, and soft skills of applicants, as well as to support jobseekers in broadening their occupational job search, increase their self-awareness and self-confidence and support targeted learning and mobility between jobs [1].

Several skills-based profiling tools have been developed across Europe, to enable diagnosing jobseekers' strengths and weaknesses in regard to personal career planning, as well as the transferability of their skills to new occupational choices and transitional labour markets, (re)integration and matching with existing employment opportunities and job vacancies. Psychometric assessment tools on the one hand and game-based assessment on the other hand have been the main methods used so far from employers in the process of personnel selection, having both strengths and limitations as well. For example, questionnaires are typically quite lengthy and due to certain obstacles, such as lack of conceptual understanding, makes the selection process cumbersome, stressful and time consuming.

This paper presents a novel approach in skill assessment developed under NADINE H2020 project. AI technology is utilised to facilitate the indirect skill assessment via agnostic content serious games, based on the evaluation of an individual's performance in playing a game outside a situational judgement test (SJT) context. The aim is to propose a different skill assessment methodology helping to overcome the multiple limitations of existing traditional skill assessment tools.

## EXISTING SKILLS ASSESSMENT METHODS

### a. Psychometric methods and tools for skill assessment

According to Wayne [2] several companies have added specific tests to their hiring processes, evaluating factors such as aptitude, personality, skills and motivation of potential employees and allowing managers to better choose the candidate, as this method can predict the individual's performance in a job.

In the area of recruitment, psychometric tests are designed to measure a person's suitability for a role based on the required personality characteristics and aptitudes, identifying the extent to which a job candidate's personality and abilities match those required to perform the role. They provide an overview of a potential employee's strengths and limitations, and this makes

it easier for the employer to understand if he/she is the right person for the job. The two basic types of psychometric assessment are **aptitude tests** and **personality tests**.

Aptitude tests are the ones that determine to what extent a candidate possesses a particular skill or capability, attempting to predict an individual's ability to succeed in an intellectual or physical endeavor. According to Vargas [3] "they are a structured tool that covers a wide variety of fields of work, some of which are more applicable, such as clinical, organizational and educational psychology, among others". These tests are structured instruments, i.e., people have to choose from the alternatives the answer that is possible according to their criteria or that fits better in their particular case [4]. The psychometric tests aim to achieve a wider information of the subject, which will allow to have more knowledge about the subject. The results obtained from the intelligence and aptitude tests, are used to measure a person's cognitive abilities and mental, emotional, developmental, and social characteristics and can be good predictors of the future performance of the individual in different situations according to González [5].

While aptitude tests predict whether a candidate will be able to perform the job, personality tests tell us whether the candidate will want to do the job and will be happy in the role. This type of psychometric evaluation is mainly oriented to know whether or not a person can adapt to a new environment, whether or not they can adapt to change or they have good interpersonal relationships and are capable of performing a given task efficiently, under stress and mainly if they can look for solutions without being startled or mentally affected [4].

### ***Advantages and limitations***

Employers use the information collected from the psychometric tests to identify the hidden aspects of persons and figure out if the person in question possesses a series of these important skills, elements that are difficult to extract from a face-to-face interview or any other selection process that contains human contact [6]. As it is common knowledge that people tend to portray themselves differently from what is actually true, in order to achieve a realistic mapping of a person's abilities, characteristics and skills for different situations, psychometric tests are the most efficient tool according to research, to discover whether a person can fit into a specific environment and act according to societal rules and values [7]. Additionally, several studies and meta-analyses support not only the validity of cognitive ability and personality tests but also their effective combination in predicting job performance [8]. The effectiveness of the tests provides a degree of satisfaction in the company since it will recruit the staff with greater ability and their skills are consistent with the functions required to perform the functions required in the position to meet the objectives [9].

Furthermore, standardised psychometric assessments constitute a scientific method widely used to *objectively* measure individuals' mental abilities, skills, intelligence, personality traits, motivations, interests and behavioural style, as they are designed, administered, and scored in a standard, or consistent, manner. A standardized test requires all test takers to answer the same questions, or a selection of questions from a common bank of questions, in the same way, and is scored in a "standard" or consistent manner, which makes it possible to compare the relative performance of individuals. Standardized tests can be administered to large populations of the same age or region, or country, and results can be compared across individuals and groups of the similar characteristics.

While standardized tests may come in a variety of forms, multiple-choice and true-false formats are widely used for large-scale testing situations because computers can score them quickly, consistently, and inexpensively. Standardized tests are considered to be a fair and objective method of assessment mainly because the standardized format, coupled with computerized scoring, reduces the potential for favoritism, bias, or subjective evaluations.

Concerning different assessment formats, while standardized tests were traditionally presented on paper and completed using pencils, and many still are, they are nowadays increasingly computer-based being administered on computers connected to online programs. Paper-based methods of assessment are widely used, where the person being tested completes written questionnaires, or writes narrative accounts of his/her progress or several skills.

Although this is appropriate for some target groups, there are circumstances in which the use of other assessment formats may be more suitable. Persons with numeracy, literacy or learning difficulties may experience problems in undertaking written assessments. One solution may be to work in conjunction with clients, using other media that is perhaps more appropriate to their needs, using ICT-based assessment tools [10].

While there is a large body of research which indicates general mental ability and personality tests as important predictors of performance, however, traditional selection methods, such as personality tests, predict job performance to some extent, whereas, they are prone to faking and social desirability [11]; [12]. There is one more issue that pertains to negative reactions to psychometric assessments. Hausknecht *et al.* [13] found psychological assessments are less accepted among applicants, which is concerning because psychological assessments tend to more strongly predict job performance, or due to fatigue that is cause of long assessment tools. Even more, traditional measurement methods, like rating scales, tests are often criticized as boring and repetitive [14], while the process of their adaptation to meet the needs of different target groups and cultures takes long.

### **b. Situational judgement tests (SJTs)**

Situational judgment tests (SJTs) have been proposed as one of the complements or alternatives to the traditional self-report questionnaires. SJTs were developed to assess an applicant's judgment regarding a situation encountered in the work place [15]. SJT items present respondents with work-related situations and various response options to deal with the situations [16].

Most SJTs are in form of a written test because the scenarios are given in a written format and applicants are asked to indicate the appropriate response alternative. The video-based SJTs consist of a number of video scenarios that describe the way a person handling a typical job-related situation [17].

### ***Advantages and limitations***

Among the strengths of SJTs are that they show criterion-related validity and incremental validity over cognitive ability and personality, while they have less adverse impact towards minorities than cognitive ability tests. Furthermore, applicant reactions towards SJTs are positive due to the job-relatedness of SJTs and they can also be used to test large groups of applicants at once [18].

Although SJTs have been advanced as alternative method for assessing skills, they have also several limitations. Specifically, SJTs might be prone to faking, practice, and coaching. In addition, the contextualized nature of SJT items makes them particularly prone to cultural differences as the culture where one lives, influences the interpretation of events and defines appropriate behaviours [19]; [20]. Furthermore, as they are context-specific instruments, it is necessary to develop different SJTs for assessing skills required in specific jobs or job families [18].

### **c. Game based skills assessment (GBA)**

In recent years, technology in the employment testing arena has proliferated and diversified in ways that have resulted in radically different methods for evaluating candidates' qualifications. It has been recently acknowledged that gamification, that is the use of game elements in non-game contexts, may be used as a replacement or supplement to traditional performance assessments. A serious game or applied game is a game designed for a primary purpose other than pure entertainment, generally referring to video games and sharing aspects of simulation but explicitly emphasizing the added pedagogical value of fun and competition. It is used by industries such as defence, education, scientific exploration, health care, emergency management, city planning, engineering, and politics [21]. Serious games are a subgenre of serious storytelling, where storytelling is applied "outside the context of entertainment, where the narration progresses as a sequence of patterns impressive in quality ... and is part of a thoughtful progress" [22].

Game-based assessment (GBA) specifically refers to methods that include *multiple* game elements and measure an individual's abilities [23]. GBAs in an organizational context can measure an array of constructs, including cognitive ability, personality, and other well-established predictors of job performance [24] [25]. Gamified assessment takes the science of psychometric tests and applies gaming elements to make the tests more engaging and more attractive to test takers and offer greater information to test users. Elements of gaming might include progressing through levels, earning points, or getting badges to create a more engaging and contemporary looking online test. The result is a psychometric test which works like a traditional psychometric test and measures the abilities it needs to measure – but includes some gaming components. Results of the gamified assessment indicate candidate competencies, skills and personality characteristics.

The application of gamification in employee testing might restrain increasing thus the assessment's predictive validity and utility in practise, while on the other hand, the advent of technology has given new perspectives to traditional selection methods, leading to more technologically advanced methods capable to reduce the cost of hiring and improve applicant reactions [12]. Consequently, serious games and gamification have already gain a place in major areas of Work/Organizational Psychology and Human Resources Management (HRM), that are recruitment, selection, training, and performance management [26].

A few skills assessment serious games are already in place (e.g. Game-Based Assessment – GBA developed by Owiwi, Simformer Business Simulation - SBS, SmartPredict, Underground), the majority of which are deeply connected to the content domain [27]. They are mainly situational games, meaning that the game mechanics were designed to reflect the features of the specific assessment content (i.e., well-defined scenarios).

***Advantages and limitations***

Serious games and gamified assessment are thought to obtain higher quality assessment data for the following reasons. Serious game performance may be more difficult for test-takers to fake in an effort to maximize their chances to be hired because desirable behaviours within the serious game may be less obvious to players. For example, personality traits might be assessed indirectly via game play behaviours, such behaviours may be less susceptible to social desirability bias, and due to the ambiguous nature of their measurement, may also be more difficult for test-takers to manipulate purposefully, as well as serious games may be better able to elicit behaviours than traditional questionnaire-based assessments. On the other hand, serious games may also elicit job-relevant behaviour more readily than is possible with questionnaires. Non-cognitive survey-based assessment measures (e.g. personality surveys, interest inventories) ask that job applicants reflect upon themselves and respond based upon their judgment about their own capabilities. The use of situational judgment tests, requiring test-takers to predict their future behaviour rather than reflect upon their existing psychological traits and states, contributed to the efforts for minimising the effects of inaccuracy introduced by this reflection while maintaining their otherwise strong psychometric properties. Nikolaou et al. [12] report that “in a gamified Situational Judgement Test (SJT) the clothing of the scenarios and answers with game elements might make the desirable behaviours less obvious to candidates and as a result, more difficult to distort intentionally or unintentionally what their reactions would be in a given situation as it is away from real life situations”. In addition, a gamified assessment environment might distract candidates from the fact that they are assessed, reducing test anxiety and promoting behaviours that are more likely to appear unconsciously instead of the desirable or socially acceptable ones [12]. Serious games as assessment tools may show similar benefits and can be used to observe and measure valuable predictor constructs, while better prediction of future work behaviour may be possible within the context of a serious game, than by using survey-based measures alone.

Another advantage of using serious games as assessment tools is their inherent ability to take advantage of current technologies. Serious games in digital (as opposed to non-digital forms such as board or card games) software environments can be programmed to capture, store, and share massive amounts of user data over time through advanced systems and innovative technologies that go beyond traditional “one-shot” assessment measures [28]. Building on their technological capabilities to capture user “behaviours,” serious games can be further programmed using artificial intelligence approaches to act as virtual expert assessors of user behaviours and skills. Because they can be programmed to evaluate “behaviours” in a more objective way, they may provide more unbiased, reliable, and precise assessments than “human” test administrators in traditional testing contexts [29]. In addition, because games can provide surrogate “digital” expertise in place of expert “human” assessors, serious games are being explored and evaluated as a means to provide more valid assessments of these “real world” outcomes in ways that could be more cost effective compared to traditional assessment approaches such as pen and paper questionnaires and direct assessment by an expert [30]. Also, gamification might increase engagement levels which in turn might lead to retention and motivation during the process of selection as well as better predictions about person-job fit [31]. Serious games provide a context for measuring and assessing a broader range of skills and constructs compared to traditional assessment approaches. These skills and constructs include competencies that have been identified in various domains as important for success in the “real

world.” Serious games are amenable to assessing these competencies because they can be designed to simulate real world environments or take advantage of fictional and fantasy contexts to elicit behaviours that demonstrate creativity, decision-making, teamwork, and leadership that can be assessed through the paths that are taken, how resources are managed and the options chosen. Users’ performance can be tracked, summarized, and then validated against other validated measures or similar behaviours in the real world [32].

Regardless the new opportunities offered by gamifying assessment, there are several conditions that still should be met, related to ensuring the psychometric properties of the game based assessment procedures, referring to reliability, consistency of measurement, validity, and fairness as described by the internationally accepted seminal documents on test validation, providing specific guidelines for the creation of psychometrically valid assessment tools. That is, serious games should produce similar scores for applicants regardless of their past experience (or lack thereof) with the game, as the number of times a person completes a personality measure not affect their personality score [33]. Also, ensuring validity measuring only the constructs intended to be measured and predicting job performance adequately is a major concern for serious game development, because most serious games are intended to elicit a wide range of highly complex skills. Whereas a psychological measure can be designed to assess a single personality trait, designing serious games to measure a single construct may be more difficult and is contrary to the typical game design process, which emphasizes a variety of interesting tasks to maximize player engagement [34]. Another concern relates to the possible unintentional discrimination against members of groups involving race, sex, religion, national origin, colour, disability status, and age. Furthermore, situational games may not be the appropriate method for assessment of different cultural populations and as content specific mechanics cannot be adjusted to different domains, there is an inherent need for different sets of mechanics to be built for each game to assess different skills in different domains [35] [36].

To sum up, while gamification is an appealing concept for employers as it potentially offers a fresh approach to selecting job candidates, the existing game based tools are mainly situational games, meaning that the game mechanics were designed to reflect the features of the specific assessment content (i.e., well-defined scenarios) and cannot be adjusted to different domains (i.e., a variety of scenarios). According to Kankaraš [37] the situational judgment techniques suffer from limitations, such as respondents’ tendency to provide socially desirable answers, or some given situations/scenarios are less appropriate for different populations or cultural groups and settings. Furthermore, the effectiveness of the use of game-based assessments in employee selection still needs to be scientifically further researched by examining the criterion related validity of a game-based assessment measuring soft skills.

### **A NOVEL METHOD USING (AI) TECHNOLOGY TO FACILITATE CONTENT-AGNOSTIC GAME BASED ASSESSMENT**

Taking into account the limitations of the existing traditional and game based assessment tools, within NADINE project, we decided to utilize content-agnostic game mechanics using artificial intelligence (AI) technology in order to assess a set of skills indirectly taking care also for effectiveness of the tools using the criterion related validity method. Our initial hypothesis was that a game provides a controlled environment with a clearly defined goal and a set of actions through which an individual can implicitly exhibit his/her proficiency of certain skills.

Through not for purpose games, which are familiar to our target group and easily accessed by different devices, our methodology focus on the indirect skill assessment based on the performance of a person, gathering data on the way a person interacts with a game (i.e., plays the game) and infer on this person's skills. This indirect skill assessment method will be grounded by the psychometric results of valid and standardized skill assessment questionnaires using Deep Learning Techniques.

In order to test the effectiveness of the indirect skill assessment tools, the participants in the research were guided to play two serious games and respond to a set of standardised skill assessment questionnaires evaluating 35 skills. The aim of this process was to combine the skills scoring acquired through the questionnaires to the way the participants play the games and in this way to export the skills scoring indirectly through the games. The analysis of the results will give an overview of the games effectiveness in skills evaluation.

### **Method**

To perform indirect skill assessment, the NADINE platform uses an artificial neural network (ANN) to perform skill assessment by evaluating the participant performance on a game. An ANN is a directed network of layers of artificial neurons that simulates the learning processes found in biological organisms (ref 10.1073/pnas.79.8.2554). At each layer, the input information is processed by the neurons and passed down to the next layers, using the weighted connections between them. During the training process, the network is fed with a set of training input data (scoring from questionnaires and moves a player makes while playing the game), computes its output and has its connection weights adjusted based on the distance between the proposed output (scoring from games) and the ground truth (scoring from questionnaires). Through this iterative process, the network error is minimized, after which point the model can be used to make predictions for data outside the training set.

The proposed model uses a pre-trained I3D network for pattern recognition (ref arXiv:1705.07750). The I3D network is a neural network that can be trained on videos to recognize patterns that can be used to predict the desired output. Instead of training the network from scratch, a process that would require a large amount of data and significant amounts of time, an already trained version of the model, pre-trained on action recognition data is used. In that state, the network is optimized to recognize patterns defining a set of 400 actions from video data. For example, the network can analyze a video and produce patterns that predict whether the video contains a person celebrating, hitting a baseball, tap dancing or any other of the 397 actions that it is trained on. By modifying only the output layer of the network, the previously learned patterns can be utilized not to identify action classes, but to predict skill value estimations on a scale of 1-10. Even though the original network is trained on a different problem domain, the low level pattern extraction performed at the intermediate layers can be utilized to identify patterns in a game session, which in turn are used for the skill assessment. During the training process, the neural network is provided with game sessions as a video with the frames containing the player's moves in a sequence, and corresponding ground truth skill values from the questionnaire scores. The network uses these training examples to learn how the pre-trained pattern recognition segment can be used to predict its new output.

A separate network is trained for each game. The final model is trained for 200 epochs on a NVIDIA Tesla T4 GPU using an 80/20 training/evaluation split, that means that the data



collected from the 80% of our sample were used for the training of the algorithms (the matching between questionnaires scoring and players' moves), while the rest 20% were used in order to evaluate the results produced from games in comparison to the questionnaires scores.

### **Participants and measures**

290 participants completed the NADINE skill assessment questionnaires. 239 individuals played the Tetris game and 216 played the 2048 game. Each participant was given instructions about each game's rules and had time to practice with the controls before submitting a formal game session. They were also assisted during the questionnaires' administration by trained facilitators.

NADINE skill assessment tool consists of 13 standardised time limited questionnaires assessing 35 hard and soft skills. The skills that are assessed through the questionnaires are:

**Interpersonal skills:** Social interaction, Team working, Intercultural competence, Coping with authority, Conscientiousness, Extraversion. **Organizational skills:** Work efficiency, Time management, Independent work, Organization. **Personal skills:** Adaptability, Reliability, Willingness to learn, Stress tolerance. **Entrepreneurial skills:** Creativity, Managerial skills, Initiative, Risk tolerance, Leadership. **Cognitive skills:** Observation, Decision making, Problem solving. **Numeracy, Mechanical skills, ICT skills, Technical skills:** Dexterity, Spatial ability, Accuracy, Language learning aptitude. The questionnaires are of different formats, including multiple choice questions with two or more options, or Likert point scales. The results are provided in a scale 1-10.

As already mentioned in our study we used two content agnostic games: Tetris and 2048 that were selected due to the fact that are user-friendly, easy to play, with non-verbal interfaces, pleasant and easily accessed by different devices, i.e. mobile phones, tablets, etc.

Specifically, the Tetris game employs a content-agnostic design and can be categorized as an action space game characterized by simple rules and a simplistic design. The game rules are evident and clear for the player to identify (i.e., the player is aware of all possible actions such as rotating and moving blocks, or accelerating their descent) while the game offers the player a fully observable action space with no other possible interaction besides going to the menu and ending the game [38].

A similar design is adopted by 2048, the online puzzle game. The game is played on a four-by-four square grid with some of the boxes filled with numbers in powers of two (2). The objective is to order the randomly generated boxes by merging or sliding them into place so as to create a tile labelled 2048 [39].

For Tetris, the collected data was a stream of board frames, each depicting the board state after a block was placed. For 2048, each session produced a board frame for every move made. The average Tetris session contained 49 frames, while the average 2048 session had 243 frames.

## **RESULTS**

After algorithms training, each skill was evaluated independently using its mean absolute error across all samples in the evaluation sets. The network performance was evaluated using a set

of 58 participants for Tetris and 48 participants for 2048 (using 20% of the data set for evaluation as indicated above) using mean absolute error for each skill. The results appear on the table 1, where the mean absolute error (MAE) reports the average distance of predicted value from the questionnaire assessment value, which is considered the ground truth. For instance, for the interpersonal skills, the 2048 model will be about 0.77 points off on average in either direction from the questionnaire assessment.

**Table 1: Mean absolute error per skill, compared to questionnaire results**

Skill	Tetris	2048
Interpersonal skills	1.22	0.77
Social interaction	1.6	1.4
Team working	1.6	1.39
Intercultural competence	1.61	1.51
Coping with authority	1.73	1.28
Conscientiousness	1.82	1.23
Extraversion	1.5	1.18
Organizational skills	1.32	1.29
Work efficiency	1.72	1.54
Time management	1.54	1.53
Independent work	1.52	1.34
Organization	1.44	1.31
Personal skills	1.28	0.74
Adaptability	1.71	1.77
Reliability	1.63	1.55
Willingness to learn	1.51	1.36
Stress tolerance	1.82	1.42
Entrepreneurial	1.32	1.13
Creativity	1.48	1.57
Managerial	1.85	1.52
Initiative	1.67	1.4
Risk tolerance	1.56	1.89
Leadership	1.76	1.68
Cognitive skills	1.48	1.66
Observation	1.79	1.95
Decision making	1.81	1.85
Problem solving	2.09	2.09
Numeracy	1.44	1.37
Mechanical skills	1.91	1.65
ICT skills	2.24	2.32
Technical skills	1.78	1.72
Dexterity	2.35	2.56
Spatial ability	2.41	2.07
Accuracy	1.8	1.89
Language learning aptitude	1.74	1.88

The models for both games have the ability to make good predictions across all skills, with the highest error having a value of 2.41 (spatial ability) for Tetris and 2.56 (dexterity) for 2048. Moreover, the average error for Tetris is 1.68, while for 2048 is 1.56. These results are quite reasonable, considering the size of the datasets.

Comparing the two models, we come to the conclusion that predictions using the 2048 data are more accurate on 24 out of 35 skills. While this may seem strange at first, taking into account the complexity of the boards of the two games and the fact that there are more training data available for the Tetris game, it can be easily explained by observing the average length of each game. Tetris games are generally shorter, with the average Tetris session having only 49 frames, while 2048 sessions tend to go on for much longer periods of time, with an average of 243 frames.

However, it is possible that by increasing the size of the Tetris dataset or employing techniques such as using board data just from the most active sub region of the game, we can bridge the gap in performance.

## DISCUSSION

Skill assessment plays an important role in personnel selection and well-validated assessment tools can be a great resource in streamlining the process. The need for developing user friendly innovative assessment tools, such as serious games, which are easy to play, easily accessed by different devices and are effective in avoiding fatigue or anxiety towards testing becomes more and more apparent. Furthermore, in times of continuous changes, such as technology advancement, organizations should not fall behind from adopting workplace trends, such as gamification. Taking also into account that the generation entering currently the workplace is highly familiar with video games and technology compared to previous generations [40] [41], organizations may supplement or replace their traditional selections methods with gamified assessment methods which are likely to increase applicants' attraction to organization and in turn, their recommendation intentions [42]. This study focused on an innovative gamified method, the indirect assessment of soft and hard skills via agnostic content serious games, such as Tetris and 2048. Through not for purpose games, NADINE main aim is to promote the accessibility and user-friendliness for the end-users, as well as to simplify the skill assessment procedure, taking into account the specific limitations related to language barriers and other difficulties, such as literacy and learning disabilities, providing a cost effective, and less time consuming screening procedure of job seekers.

Key results of the study revealed that the trained neural models for both games have strong skill assessment capabilities. As indicated by the low error rate across most skills, there is a considerable degree of correlation between an individual's move patterns on the two games and his/her skill values. Comparing the two models, even though the Tetris game has a larger set of available actions per frame, the 2048 models were able to provide better performance in 24 out of the 35 skills, due to the higher number of moves per session, which allows for more intricate sequences of patterns to occur and be detected by the model. It could also be argued that in the 2048 game, providing the player with a goal, makes him/her more likely to exhibit his/her skills by developing a long term strategy, working towards that goal. Tetris players are generally forced to make short term plans as they are affected by the time pressure and the lack of knowledge of the next sequence of blocks, which translates to many of them converging to greedy strategies.

Summing up, using the games Tetris and 2048 we proved that there is indeed correlation between a person's action sequence and his/her different skills, by creating our own novel game/skill vector datasets, training and evaluating on a separate sample. It came out that for

most skills we can expect estimations that are close to the real skills scoring acquired through the psychometric questionnaires, especially when using a game with a high number of actions per session, like in the case of 2048. These initial results are very encouraging in implementing in personnel selection process the indirect skill assessment via games use, under the condition that for the algorithms training are used valid and standardized psychometric tools. By using content agnostic games, we can assist employers in making a first screening among big numbers of candidates within limited time, as using such a method eliminates the need for human supervision and significantly reduces the assessment time, and allows for skill assessment at scale.

Furthermore, content-agnostic designs can be used across different games as they do not rely on visual representations, are independent of game mechanics [43], [44] and can foster tasks that are situated in a multitude of domain assessments while allowing for a link between player behavior and skill assessment [45]. This new form of assessment includes online games and the use of large amounts of big data pulled from many sources, including test responses, test-taking behavior, applications. Speedy processing, lower costs, convenient access, and applicant engagement are often and rightfully cited as the practical advantages for using this kind of selection tool [46]. Also, as the design of the content-agnostic games is simplistic, independent of culture-specific content, and features rules that are easy to follow, the games are not in need of adaptation of any kind, and are not expected to be problematic for culturally different populations in terms of understanding the game objectives or the playing rules, while they can easily be accessed by different devices, i.e mobile phones, tablets, etc.

Therefore, indirect assessment through gamification can play here an important role in the employee selection process, as an alternative or supplement to the traditional selection methods, such as psychometric tests, situational judgement tests (SJTs). Using content agnostic games, we might reduce faking and socially desirable responses, since these may be less obvious while completing a “game”, and as a result, the quality of information about candidates and the prediction of job performance may increase [47]. Moreover, engaging gamification in the selection process may make the process more interactive and attractive to candidates, as the use of game elements might promote fun, challenge and interaction [42].

Even though the initial results are encouraging, more research is needed to establish the predictive validity of gamified indirect skill assessment above and beyond what traditional selection methods evaluate. If the training data was scaled up to even a few thousand entries, it is very possible that the resulting models would be able to perform skill assessment with a minimal error on most skills. Therefore, there is still need for additional data collection and engagement in the survey of more participants, so as to further training the algorithms with more data. Additionally, more games could be tested in the same way as Tetris and 2048, so as to have a pool of tools that could be used in indirect skill assessment and to increase the matching between psychometric skills assessment results and indirect skills assessment results.

Though, given the increasing interest in cost effective and valid assessment tools in recruitment process, we hope that this study will stimulate further empirical evidence on indirect gamified assessment methods and their potential usefulness.

## References

- [1]. Blázquez M. (2014). *Skills-Based Profiling and Matching in PES*. Luxembourg: Publications Office of the European Union.
- [2]. Wayne, M. (2010). *Human Resources Administration - 11th Edition*. Mexico: Pearson Education.
- [3]. Vargas, L. F. (2016). Construcción de pruebas psicométricas: aplicaciones a las ciencias sociales y de la salud Livia, J. & Ortiz, M. (2014) Lima: Editorial Universitaria. 166 p. *Revista Digital de Investigaciones Docencia Universitaria*, 10(2), 92-94.
- [4]. Chipana-Castillo, C., Miranda-Roca, GJ. & Vicente-Ramos, W. (2021). Effectiveness of psychometric tests for the selection of personnel in jobs in the retail sector, *Management Science Letters* 11, 1583–1590.
- [5]. González, F. (2007). *Psychological evaluation instruments*. Havana: Medical Sciences.
- [6]. Andrews, J. & Higson, H. (2008) Graduate Employability, 'Soft Skills' Versus 'Hard' Business Knowledge: A European Study, *Higher Education in Europe*, 33:4, 411-422 <https://doi.org/10.1080/03797720802522627>
- [7]. Schulz, B. (2008). The importance of soft skills: Education beyond academic knowledge. *Nawa Journal of Communication*, 2(1), 146-154.
- [8]. Schmitt N. (2014). Personality and cognitive ability as predictors of effective performance at work. *Annual Review of Organizational Psychology and Organizational Behavior*, 1(1), 45–65. <https://doi.org/10.1146/annurev-orgpsych-031413-091255>
- [9]. Ferro, R. & Vives, C. (2004). An analysis of the concepts of effectiveness, efficacy and efficiency in psychology, *Tremédica*, 5 (16). 97-99.
- [10]. Dewson S., Eccles J., Tackey N.J., Jackson A., (2000). *Guide to Measuring Soft Outcomes and Distance Travelled*. UK: The Institute for Employment Studies, Department for Education and Employment, Crown. Available at <http://www.employabilityinscotland.com/media/83581/guide-to-measuring-soft-outcomes-distance-travelled.pdf>
- [11]. Morgeson, F., Campion, M., Dipboye, R., Hollenbeck, J., Murhpy, K., & Schmitt, N. (2007). Are we getting fooled again? coming to terms with limitations in the use of personality tests for personnel selection, *Personnel Psychology*, 60, 1029-1049. <https://doi.org/10.1111/j.1744-6570.2007.00100.x>
- [12]. Nikolaou, I., Georgiou, K. & Kotsasarlidou, V. (2019). Exploring the Relationship of a Gamified Assessment with Performance, *The Spanish Journal of Psychology*, 22. <https://doi.org/10.1017/sjp.2019.5>.
- [13]. Hausknecht J., Day D. & Thomas S. (2004). Applicant Reactions to Selection Procedures: An Updated Model and Meta-Analysis, *Personnel Psychology*, 57. <https://doi.org/10.1111/j.1744-6570.2004.00003>
- [14]. Sinickas, A. (2007). Finding a cure for survey fatigue, *Strategic Communication Management*, 11(2), 11.
- [15]. Weekley, J.A. and Ployhart, R.E. (2006). *Situational Judgment Tests: Theory, Measurement and Application*, Jossey-Bass, San Francisco, CA.
- [16]. Herde, C. N., Lievens, F., Solberg, E. G., Harbaugh, J. L., Strong, M. H., & Burkholder, G. J. (2019). Situational judgment tests as measures of 21st century skills: Evidence across Europe and Latin America. *Journal of Work and Organizational Psychology*, 35(2), 65–74.
- [17]. Weekley, J.A. and Jones, C. (1997). Video-based situational testing, *Personnel Psychology*, Vol. 50, No. 1, pp. 25-49.

- [18]. Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: A review of recent research. *Personnel Review*, 37(4), 426–441. <https://doi.org/10.1108/00483480810877598>
- [19]. Heine, S. J. & Buchtel, E. E. (2009). Personality: The universal and the culturally specific. *Annual Review of Psychology*, 60, 369-394. <https://doi.org/10.1146/annurev.psych.60.110707.163655>
- [20]. Lytle, A. L., Brett, J. M., Barsness, Z. I., Tinsley, C. H., & Janssens, M. (1995). A paradigm for confirmatory cross-cultural research in organizational behavior. *Research in Organizational Behavior*, 17, 167-214.
- [21]. Djaouti, D., Alvarez, J. & Jessel, J.P. (2011). Classifying Serious Games: the G/P/S model, Handbook of Research on Improving Learning and Motivation through Educational Games: Multidisciplinary Approaches. IGI Global.
- [22]. Lugmayr, A., Suhonen, J., Hlavacs, H., Montero, C., Suutinen, E. & Sedano, C. (2016). "Serious storytelling - a first definition and review". *Multimedia Tools and Applications*. 76 (14): 15707–15733. <https://doi.org/10.1007/s11042-016-3865-5>.
- [23]. Bhatia, S. (2018). *Applicant Reactions to Game-Based Assessments: The Effects of Flow, Fairness, and Fit*, MI: Michigan State University, Lansing.
- [24]. Ellison, L.J., McClure Johnson, T., Tomczak, D., Siemsen, A. and Gonzalez, M.F. (2020), "Game on! Exploring reactions to game-based selection assessments", *Journal of Managerial Psychology*, Vol. 35 No. 4, pp. 241-254. <https://doi.org/10.1108/JMP-09-2018-0414>
- [25]. Glover, I. (2013). Play as you learn: Gamification as a technique for motivating learners. In J. Herrington et al. (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications* (pp. 1999-2008). VA: AACE.
- [26]. DuVernet, A. M., & Popp, E. (2014). Gamification of workplace practices. *The Industrial-Organizational Psychologist*, 52, 39–44.
- [27]. Van Eck, R. (2006). Digital game-based learning: It's not just the digital natives who are restless. *Educause Review*, 41, 16. Retrieved from <https://www.researchgate.net/>
- [28]. de Klerk, S., Veldkamp, B. P., & Eggen, T. J. (2015). Psychometric analysis of the performance data of simulation-based assessment: A systematic review and a Bayesian network example, *Computers & Education*, 85, 23–34. <https://doi.org/10.1016/j.compedu.2014.12.020>
- [29]. Bellotti, F., Kapralos, B., Lee, K., Pablo Moreno-Ger, & Ber-ta, R. (2013). Assessment in and of serious games: An overview. *Advances in Human-Computer Inter-action*, 2013 <https://doi.org/10.1155/2013/136864>.
- [30]. Mislevy, R. J., Corrigan, S., Oranje, A., DiCerbo, K., Bauer, M. I., von Davier, A., Drasgow, F. (2016). Psychometrics and game-based assessment. In F. Drasgow (Eds.), *Technology and Testing: Improving Educational and Psychological Measurement* (pp. 23–48). New York, NY: Routledge.
- [31]. Chamorro-Premuzic T., Akhtar R., Winsborough D., & Sherman R. A. (2017). The datafication of talent: How technology is advancing the science of human potential at work. *Current Opinion in Behavioral Sciences*, 18, 13–16. <https://doi.org/10.1016/j.cobeha.2017.04.007>.
- [32]. Kato & de Klerk, 2017) Kato P. M. & de Klerk S., (2017) Serious Games for Assessment: Welcome to the Jungle, *Journal of Applied Testing Technology*, Vol 18(S1), 1-6, 2017.
- [33]. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests, *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555>

- [34]. Schell, J. (2008). *The art of game design*. Burlington, MA: Morgan Kaufman.
- [35]. Baron, T., Heath, C. & Amresh, A. (2016). Towards a context agnostic platform for design and assessment of educational games. In Connolly T. & Boyle D. (Eds.), *Proceedings of the European Conference on Games Based Learning* (pp. 34-40). Scotland, GB: The University of West Scotland.
- [36]. Tremblay, J. & Verbrugge, C. (Eds.). (2015). A game genre agnostic framework for game-design. *IEEE/ACM 4<sup>th</sup> International Workshop on Games and Software Engineering, Florence, Italy*. <https://doi.org/10.1109/GAS.2015.9>
- [37]. Kankaraš, M. (2017). Personality matters: Relevance and assessment of personality characteristics, *OECD Education Working Papers*, No. 157, OECD Publishing, Paris. <http://dx.doi.org/10.1787/8a294376-en>
- [38]. Stiegler, A., & Zimmermann, G. (2015). Gamification and accessibility. In J. Zhou and G. Salvendy (Eds.), *Human Aspects of IT for the Aged Population: Design for Aging* (pp. 145-154). [https://doi.org/10.1007/978-3-319-20892-3\\_15](https://doi.org/10.1007/978-3-319-20892-3_15)
- [39]. Neller, T. (2015). Pedagogical possibilities for the 2048 puzzle game. *The Journal of Computing Sciences in Colleges*, 30, 38-46. Retrieved from <https://cupola.gettysburg.edu/>
- [40]. Bannon, S., Ford, K., & Meltzer, L. (2011). Understanding millennials in the workplace. *The CPA Journal*, 81(11), 61–65. Retrieved from: <https://www.nysscpa.org/news/publications/the-cpa-journal/article-preview?ArticleID=10721>
- [41]. Gibson, L. A. & Sodeman, W. A. (2014). Millennials and technology: Addressing the communication gap in education and practice. *Organization Development Journal*, 32(4), 63–75.
- [42]. Gkorezis P., Georgiou K., Nikolaou I. & Kyriazati A. (2020): Gamified or traditional situational judgement test? A moderated mediation model of recommendation intentions via organizational attractiveness, *European Journal of Work and Organizational Psychology*, <https://doi.org/10.1080/1359432X.2020.1746827>
- [43]. Cutumisu, M., Blair, K., Chin, D., & Schwartz, D. (2015). Posterlet: A game-based assessment of children’s choices to seek feedback and to revise. *Journal of Learning Analytics*, 2, 49-71. <https://doi.org/10.7939/R31R6ND7N>
- [44]. Baron, T. (2017). *An architecture for designing content agnostic game mechanics for educational burst games* (Doctoral dissertation, Arizona State University).
- [45]. Mislevy, R., Steinberg, L., & Almond, R. (2003). Focus article: On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1, 3-62. [https://doi.org/10.1207/S15366359MEA0101\\_02](https://doi.org/10.1207/S15366359MEA0101_02)
- [46]. Tippins, N. T., Oswald, F. L., & McPhail, S. M. (2021). Scientific, Legal, and Ethical Concerns about AI-Based Personnel Selection Tools: A Call to Action, *PsyArXiv*, 28. <https://doi.org/10.31234/osf.io/6gczw>
- [47]. Armstrong, M. B., Landers, R. N., & Collmus, A. B. (2015). Gamifying recruitment, selection, training, and performance management: Game-thinking in human resource management. In *Emerging Research and Trends in Gamification* (pp. 140-165). IGI Global. <https://doi.org/10.4018/978-1-4666-8651-9.ch007>.