Effectiveness of Epistemic Games as a medium to incorporate Situational Judgement Tests

Abstract—

An organisation's performance is highly dependent on its ability to attract and identify the right talent to meet its strategic objectives. A key challenges that recruiters face, is to identify a method which would be in a position to measure a wide range of non academic attributes reliably. Conducting large scale interviews can be costly and have also been criticised because of the individual bias that crops up during selection. Personality tests have limited evidence to support their predictive validity for selection purposes in high stake settings. Situational Judgement Tests (SJTs), as a multi dimensional measurement method has been gaining prominence since it has evidence of high predictive and incremental validity and more importantly has a favourable applicant perception because it measures the participants disposition and reaction to a scenario presented in an organisational context. In this paper we propose to measure the effectiveness of using epistemic games as a stimulus medium in presenting SJTs. Epistemic games provide the possibility of incorporating SJTs and other non intrusive assessments and contextualising the scenario based on the game state .

Index Terms—Situational Judgement Tests, Assessments, Simulation Games, Evidence Centred Game Design

"The competition to hire the best will increase in the years ahead. Companies that give extra flexibility to their employees will have the edge in this area."

Bill Gates

I. Changing Face of employee selection

Organisations are finding it hard to identify the right person for the right job. The general practise of identifying a candidate based on their academic performance and knowledge is not proving to be effective. Research and practise have shown evidence of non academic parameters such as empathy to fellow colleagues, teamwork and integrity are critical predictors of job performance and training outcomes[1]. Recruiters often grapple with the question of determining the tool which will enable them to access these non academic parameters reliably. Personality tests can prove to be cost prohibitive and there is limited support of their predictive validity for high stakes selection.[2]

Corporates are turning towards situational judgement tests (SJTs) to measure non academic factors in a candidate. SJTs capture a candidates response to a set of situations that might occur while working with an organisation. They are preferred for a variety of reasons, owing to their conceptual appeal as a measure of a candidate's ability to solve problems and make judgements in applied settings. SJTs also prove to have incremental validity over other established measure in ability and personality domain. They also seem to reduce the adverse impact against groups in a recruitment setting[3]

SJT as a method has been existence for several decades, but has not been widely used. These tests have become popular among practitioners over the last decade. SJTs come across as a multidimensional tests rather than unidimensional[4].

SJTs are carefully designed after performing detailed job analysis and consulting subject matter experts to arrive at the scoring pattern. The situation is either presented in a text format elaborating the scenario or presented in the form of a video. While designing SJTs, researchers predominantly use two different response format. The first one measures a candidates knowledge ("What is the best option?") and the second is to measure the behavioural pattern ("What would you most likely do?"). SJTs growing popularity is because of its face, content and predictive validity[5]. SJTs have their grounding on the behavioural consistency theory, which asserts that past behaviour is the best predictor of future behaviour[6].

The fact that the situations presented in SJTs are job related increases the job relatedness of the SJT items. However SJTs need to be presented in context to increase the fidelity with which they present the stimuli. The extent to which the scenario is consistent to what an employee might encounter in a workplace increases the fidelity of the task stimulus[7].

The stimulus is presented in the form of written tests, where applicants are asked to indicate their appropriate response and such methods have low stimulus fidelity. The second method is to provide the scenario in the form of a video or multimedia input, supported by rich visuals and narration, thereby increasing the stimulus validity[8].

To date an alternative stimulus mechanism has not been explored to administer SJTs. The authors have incorporated SJTs in a Multi Player Role Playing Business Simulation Game. The present study proposes to find the effectiveness of Epistemic games as a medium to incorporate SJTs combined with evidence based assessments in a games.

II. Situational Judgement Tests - Literature Review

The 21st century worker is expected to be a multitasker simultaneously managing tasks with great dexterity and collaborating with employees in a rapidly changing environment. Blockchain, Robotic Process Automation, Bots, Artificial Intelligence and Machine Language once words that appeared in science fiction movies have become a reality that an employee needs to deal with. In such rapidly changing environment, organisations are exploring newer methods to assess candidates beyond their academic knowledge.

As a contextualised measurement methods, SJT are becoming increasingly popular among practitioners to measure interpersonal and non academic parameters during the selection stage of an interview process. They also offer a cost effective alternative as against other personality tests.

Development of SJTs
Development of SJT is a three stage process. The researcher needs to perform a detailed job analysis noting down critical incidents of work simulations captured from the subject matter experts. The incidents are further grouped and representative

scenarios are designed [9]. As a second step, different group of subject matter experts are asked to generate one or more responses to each situation. Finally the responses are rationally or empirically scored before SJTs are administered.

Reliability
Reachers have checked the reliability and internal consistency. Studies show that the internal consistency coefficient varied between 0.43 and 0.94. Factors such as the length and the response instructions moderate the variability of internal consistency[10]. However for a multidimensional nature of the instrument, it has been studied that a test-retest reliability is a better measure of reliability and it has been found that the reliability values is 0.84 for SJTs[11].

Criterion-related validity
The most important question while choosing an instrument that aids in assessing a candidate for a job would be to measure its ability to predict job related criteria. It has been found that the corrected correlation between SJTs and job performance was 0.34 and uncorrected was 0.26. The difference was explained by moderating factors such as SJTs developed after performing job analysis and the ones that did not[10].

Incremental validity
Various researches conducted have found that SJTs have the ability to predict job performance of a candidate over cognitive ability, knowledge, job experience and personality[12]. It was also found that the SJT's correlation with job performance, cognitive ability and personality vary widely because the researcher could construct scenarios where they can contribute to a predictor composite or offer zero incremental validity[10]

Utility
SJTs can be administered across a wide range of participants over the internet and since the answers are scored a priori it is convenient and less time consuming as against other alternatives such as assessment centres.

Construct Related Validity
SJTs with knowledge instruction correlated highly with cognitive ability (0.36) and the ones with behavioural instructions correlated with Agreeableness(0.37), Conscientiousness(0.34) and Emotional Stability(0.35). It was observed that SJTs with knowledge instruction predict maximal performance measure and that with behavioural tendency measure typical performance measures.

Adverse Impact
The primary concern is to understand if there exists any group or class of individuals that would perform well in SHTs as against others. Studies have found positive impact of race and gender on SJT performance.

Applicant Perceptions
Applicants are happy to be evaluated on their ability to perform a task in a organisational context. In this regard SJTs have been viewed favourably by participants [10]. Video based formats received resulted in a higher positive perception than the written formats.

Fakability
It was found that SJTs have a low faking effects as compared to personality measures[13] and the ones with stronger cognitive loading were less fakable. Behavioural instructions were easily fakable than knowledge based instructions.

III. EPISTEMIC GAMES AS A STIMULUS MECHANISM

Situation queues are assumed to form a vital element of SJTs and hence are considered context dependent[15]. However recent research points to the direction that SJTs designed to measure the general domain knowledge of a particular trait expression of work effectiveness will be applicable across other occupations[16]. This also implies that SJTs have the power to predict behaviour across various job domains and social situations.

Digital game based environments offer a unique possibility to virtually simulate an organisational setup. Epistemic Games are designed to be pedagogical tools for the digital age where the player learns to think like professionals by playing a simulated game.This ability to simulate a professional environment becomes conducive to administer SJTs within epistemic games.

Games are inherently an assessment tool. The participants are subjected to various tasks and situations and the response captured to determine the future course of action.

Evidence-centred designs (ECD) principles used while designing epistemic games aids in creating assessment frameworks. ECD creates a framework by combining competency, evidence and task models[14].

IV GAME DESIGN

A proprietary game designed by the authors, simulates a business environment, where students don the hat of heads of functional departments such as Marketing, Operations, Finance, HR, IT and International Business. Participants take decisions pertaining to their department and the outcome of the decisions impact not only their department but other departments as well.

The game is designed using ECD principles and the actions of players are tracked by the games analytical engine.

Figure 1.0 - Game Screen

The participant chooses a virtual avatar in the game and interacts with other departments, customers, vendors and senior management. The decision that a participant takes in the game are broadly classified into two categories. The decisions taken to operationally run the department is called as operation decisions, for e.g If the participant is playing the role of an HR, they would have to record their decisions in terms of Recruitment, Training, Welfare Programs, Promotion and Retrenchment.
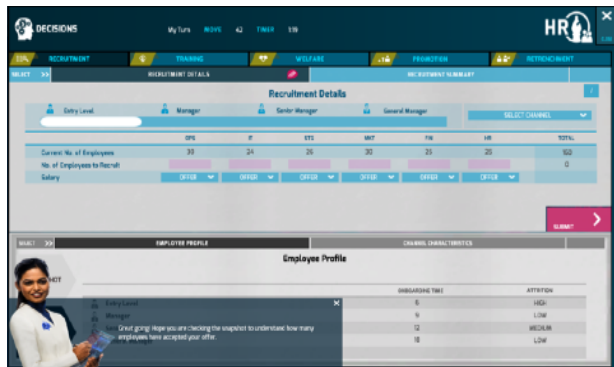

Figure 1.1 - HR Decision Console

The second form of decisions are situational judgemental decisions. SJTs are embedded in the game and participants respond to various scenarios during the game play.
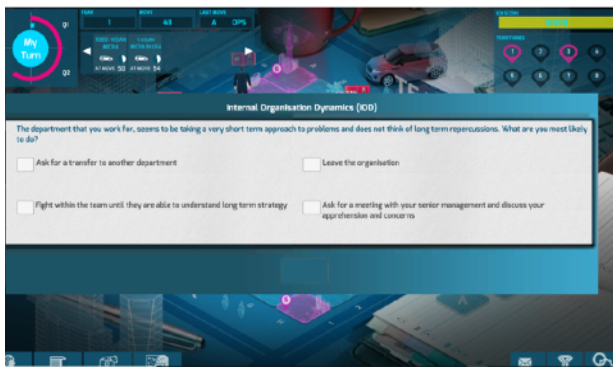

Figure 1.2 - Written SJT in the game

The scenarios are populated based on the game state and the role chosen by the participant. This method provides context to the participant. The responses to scenarios are recorded. Epistemic games provide an advantage of simulating as many scenarios as possible. Since the SJTs are embedded as constituent element of the game play, they seamlessly blend and provide a realistic view to the participant.

V. Methodology & Analysis

As a pilot study the game was administered to post graduate students who are currently pursuing their MBA in a B-School in India. A total of 120 students participated in the study. Participants were randomly assigned into teams of 6 to form a group.

A set of six dimensions were chosen after consultation with HR professionals on the dimensions that they would look for in a management candidate while recruiting. We had to carefully evaluate the parameters and choose parameters that are context independent.

The parameters chosen for different roles are different. For e.g, we choose Grievance Handling, Favouritism, Organisational Bullying, Gender Sensitivity, Personality Clashes and Discrimination as the six parameters for SJTs for an HR role. Each parameter had 5 scenarios.

After consultation with the subject matter experts in arriving at the questions and response. We tested the scenarios with a smaller sample of 15 participants. This was done outside the game, in-order to determine the face validity of the scenarios.

This design method addresses a few critical aspects in terms of Fakability, Applicant Perceptions and Utility. Participants had a lesser incentive to deliberately distort their response. Since the game had other elements as combined with SJTs and participants did not get the feeling that they were being assessed. Students and participants in general have a favourable disposition towards games and hence SJTs presented within a gaming framework was looked favourably. In terms of utility, online simulation games can be downloaded by a large numbee of participants and conducting an assessment based on SJTs can be cost effective.

The other advantage that the games have over the written and video formats, is the ease by which the scenarios can be contextualised and customised. The game that we have designed is completely configurable to handle nested scenarios, in which the response to an option will lead to another scenario depending on the option chosen. It can also handle scenarios based on video and audio inputs.

More importantly when SJTs are combined with the other decision making setups which are present in a traditional business simulation, provides the opportunity to measure aspects of the skills, knowledge, values, identity and epistemology (SKIVE) of the participant. The SKIVE parameters comprise an epistemic frame of a profession and hence provide a holistic assessment methodology.

In our preliminary analysis we found that female respondents scored better in SJTs in line with prior research[17].
The below table presents the findings of one of the 6 parameters on grievance handling

Handling grievances and finding amicable solutions is a key aspect of interpersonal skills. The responses from the pilot

group was not surprising and was in line with previous research in the area of grievance handling. The three probable actions that a manager would undertake while handling grievance,1) respond to the situation appropriately, 2) procrastinate on the decision and 3) not act on the situation.

SJT Responses - Grievance Handling

Category	Scenario - Central Theme	Male Responses	Female Responses
Grievance Handling	Shift change request from an employee you don't like	7% will not act on the situation 12% will escalate the situation 24% will evaluate the situation and resolve	9% will not act on the situation 7% will escalate the situation 30% will evaluate the situation and resolve
	Shift change request by a high performing employee which will hamper productivity.	16% will resolve the situation amicably 13% will not act 13% will procrastinate	20% will resolve the situation amicably 4% will not act 34% will procrastinate
	Shift allocation and complains of favoutitism	35% will review and act on the situation 2% will confront the person responsible 5% will not act	45% will review and act on the situation 7% will confront the person responsible 5% will not act
	Employee complaints about Food	39% will act on the complaint 11% will not act on the situation	48% will act on the complaint 2% will not act on the situation
	Grievance on promotion	22% will council and act on the grievance 24% will procrastinate	28% will council and act on the grievance 27% will procrastinate

It was found that female respondents compared to their male counterparts tend to either confront or act on the situation, while both of them equally procrastinate and male respondents compared to their female counterparts might choose not to act.

When we look at individual responses, we would be in a position to identify candidates based on the desired disposition. Also SJTs in the game also gives us the perspective of looking at aggregates responses on how most likely a person would respond and hence provides rich data for training need analysis and corrective action within the organisation.

VII DISCUSSION

SJTs are a cost effective and have high predictive validity of job performance of a candidate. The tool is gaining prominence over practitioners and organisations have started adopting it to measure non academic performance. Even though the instrument has been reported to have high incremental validity and construct related validity, its effectiveness is moderated by the stimulus medium.

The stimulus presented to the participant provides the context to the assessment tool and hence is viewed to be a critical factor in determining the stimulus validity. In this paper we have explored the effectiveness of using an alternative stimulus mechanism apart from the existing written and video formats. We have chosen to integrate SJTs in a business simulation epistemic game and have conducted a pilot study among post graduate students pursuing their MBA. Our initial analysis is in line with established research on the, perception and utility of the SJTs. The game platform offers added advantage in terms of providing other contexts to the scenarios, making it more realistic. Moreover when SJTs are connected with other evidence basesd assessments increases the measurable validity in predicting job performance.

VIII LIMITATIONS & FUTURE RESEARCH

The limitations of the current research stems from the fact that the model was validated with a smaller sample. We would like to test the validity of the model with larger audience in the coming days and validate the results.

REFERENCES
1. Prideaux D, Roberts C, Eva K, Centeno A, McCrorie P, McManus C, Patterson F, Powis D, Tekian A, Wilkinson D. Assessment for selection for the health care professions and specialty training: consensus statement and recommendations from the Ottawa 2010 conference.
2. Albanese M, Snow M, Skochelak S, Huggett K, Farrell P. Assessing personal qualities in medical school admissions. Acad Med 2003
3. Filip Lievens, Paul R Sackett & Tim Buyse, The effect of Responce Instructions on Situational Judgement Test Performance Validity in a high stakes context, Journal of applied Psychology, 2009
4. McDaniel M, Whetzel D. Situational judgement tests. In: Whetzel DL, Wheaton GR, eds. Applied Measurement: Industrial Psychology in Human Resources Management. Mahwah, NJ: Lawrence Erlbaum Associates 2007
5. Christian M, Edwards B, Bradley J. Situational judgement tests: constructs assessed and a meta-analysis of their criterion-related validities. Pers Psychol 2010;
6. Motowidlo SJ, Dunnette MD, Carter GW. An alternative selection procedure: the low-fidelity simulation. J Appl Psychol 1990;
7. Filip Lievens, Helga Peeters and Eveline Schollaert, Situational judgment tests: a review of recent research, emerald publishing, 2007
8. Funke, U. and Schuler, H. (1998), "Validity of stimulus and response components in a video test of social competence", International Journal of Selection and Assessment.
9. Bergman, M.E., Drasgow, F., Donovan, M.A. and Henning, J.B. (2006), "Scoring situational judgment tests: once you get the data, your troubles begin", International Journal of Selection and Assessment,

10. McDaniel, M.A., Hartman, N.S., Whetzel, D.L. and Grubb, W.L. (2007), "Situational judgment tests, response instructions, and validity: a meta-analysis", Personnel Psychology,

11. Bruce, M.M. and Learner, D.B. (1958), "A supervisory practices test", Personnel Psychology,

12. Chan, D. and Schmitt, N. (1997), "Video-based versus paper-and-pencil method of assessment is situational judgment tests: sub-group differences in test performance and face validity perceptions", Journal of Applied Psychology,

13. Hooper, A.C., Cullen, M.J. and Sackett, P.R. (2006), "Operational threats to the use of SJTs: faking, coaching, and retesting issues", in Weekley, J.A. and Ployhart, R.E. (Eds), Situational Judgment Tests: Theory, Measurement and Application,

14. Sandeep Rambhatla, Deepak Gupta, Evidence Centred Approach to Measuring Learning outcomes amongst Management Students using Epistemic Games, IEEE Conference, 2017

15. Krumm, S., Lievens, F., Huffmeier, J., Lipnevich, A. A., Bendels, H., & Hertel, G. (2015). How "situational" is judgment in situational judgment tests? Journal of Applied Psychology,

16. Kell, H. J., Motowidlo, S. J., Martin, M. P., Stotts, A. L., & Moreno, C. A. (2014). Testing for independent effects o prosocial knowledge and technical knowledge on skill and performance. Human Performance,

17. McDaniel, M. A., & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed International Journal of Selection and Assessment.