Constructing Cognitive Profiles for Simulation-Based Hiring Assessments

Rebecca Kantar
Imbellus
1085 Gayley Ave.
Westwood, CA 90024
rebecca@imbellus.com

Keith McNulty
McKinsey & Company
1 Jermyn St, S. James
London, UK
Keith_mcnulty@mckinsey.com

Erica L. Snow
Imbellus
1085 Gayley Ave.
Westwood, CA 90024
esnow@imbellus.com

Matthew A. Emery
Imbellus
1085 Gayley Ave.
Westwood, CA 90024
memery@imbellus.com

Richard Wainess
Imbellus
1085 Gayley Ave.
Westwood, CA 90024
rwainess@imbellus.com

Sonia D. Doshi
Imbellus
1085 Gayley Ave.
Westwood, CA 90024
sdoshi@imbellus.com

ABSTRACT

Imbellus is an assessment company that builds immersive simulation-based assessments designed to evaluate cognitive processes. The work described here explores our partnership with McKinsey & Company, a best-in-class management-consulting firm, to build a simulation-based assessment that evaluates incoming applicants' cognitive skills and abilities. Our simulation-based assessments are designed to produce a substantial amount of information about the incoming applicants, including metacognitive skills, decision-making processes, and situational awareness (to name a few of the constructs we measure). This paper will explore the rich telemetry data we collect and quantify, as well as the novel scoring and exploratory techniques we are conducting to gain insight into applicants' cognitive profiles. We will present our initial findings and describe implications of our current work for the fields of artificial intelligence, educational data mining, and assessment.

Keywords

Cognitive Assessment, Learning Science, Machine Learning

1. INTRODUCTION

Imbellus assessments are designed to provide a wealth of information concerning applicants' cognitive skills and profiles. In contrast, traditional standardized cognitive assessments primarily evaluate content mastery, processing speed, and memory. The rise of automation makes insights around domain knowledge, processing speed, and memory less relevant features of human cognition, while higher level, complex cognitive abilities become features that make all the difference in individuals' preparedness for modern work and life. Imbellus assessments evaluate what have historically been hard-to-measure skills like problem-solving, creativity, systems thinking, and critical thinking. To take a practical approach to designing good assessments, Imbellus partners with industry leaders whose employees leverage key 21st Century skills at an elite level. Our early work with McKinsey & Company, a best-in-class management-consulting firm, has involved building an assessment to gauge incoming applicants' cognitive skills and abilities, which will be used to construct profiles of each applicant.

Standardized cognitive assessments were developed in the late 1800s to "stratify students of different abilities into different curricular paths" [9]. The release of Goddard's IQ formula and the Stanford-Binet cognitive assessment in the early 1900s launched a movement of mass testing in the United States. The College Entrance Examination Board, now the College Board, was established in 1923 to define a set of college admission standards through the dissemination of the Scholastic Aptitude Test (SAT) [3]. In 1959, the American College Test (ACT) was released as an alternative to the SAT [3]. The ACT's stated goal is to "measure information taught in high school," instead of evaluating cognitive reasoning skills. [8]. The ACT and SAT set college admissions standards, which became significant shaping forces. Today over 39 Advanced Placement tests and 20 SAT Subject tests dictate the curriculum in our K-12 education system and influence infrastructure and resource allocation. The ACT and the SAT focus on standardized content in mathematics, writing, science, and other subject-specific areas to create objective metrics. [6]. While widely adopted across the nation, these assessments have

"revealed little about specific cognitive abilities or predicted performance" [3].

In response to the shortcomings in both the methodology of and substance of traditional standardized college admissions tests, employers have adopted other traditional cognitive ability or intelligence tests in an effort to glean more predictive insights on applicants' cognitive profiles. Most cognitive ability tests measure "reasoning, perception, memory, verbal and mathematical ability" [1]. These assessments, like standardized admissions tests, focus on content mastery, processing speed, and memory. These factors ignore the increasing need to develop and measure capabilities required by the 21st-century workforce. These tests ignore the cognitive process that users engage in during that task.

Past their shortcomings in predictive validity, most cognitive assessments are paper-and-pencil multiple-choice tests, a medium for evaluating cognitive skills that artificially constricts the nature of possibility spaces framing users' potential cognition. Multiple choice tests demand asking clear, static questions about some subject matter where one of n choices is right and n-1 of n choices are wrong. Such a scenario, at an abstract level, is at odds with the nature of modern demands on cognition. Traditional admissions tests focus on product scores (i.e., correctness) not the process of how (i.e., strategy) a user got there. It is vital to understand a user's cognitive process, as cognition by its nature is dynamic across time and tasks.

Beyond content irrelevance, the degree to which today's standardized admissions tests can be "gamed" leads to inequity in opportunity for success. Users who have the resources to master the testing process are more likely to perform better on the assessments. The College Board reported a substantial correlation of r=.42 between socioeconomic status and SAT scores [4]. The SAT's correlation with socioeconomic status is higher than The College Board's self-reported correlation of r=.33 between SAT score and first-year college GPA [4].

Imbellus assessments focus on evaluating how people think instead of what they know. Through our scenarios that take place in our simulation-based environments, we observe details of users' cognitive processes, not just their end choices. We've designed our assessments to discount the high value placed on memory and processing speed in traditional cognitive assessments. The simulation-based assessment discussed in this paper consists of several scenarios embedded in an abstracted natural world environment. Users interact with a series of challenges involving natural terrain, plants, and wildlife (See Figure 1). We designed each scenario as an abstract representation of the problem-solving capabilities and processes required to succeed on the job. This abstraction allows us to transpose skills to a new context with a similar structure to the first—known as far transfer [5]. We strategically chose the natural world as a setting for our tasks because it offers an accessible context for a global population.

Second, our problem-solving assessment focuses on skills mastery rooted in cognitive and learning science theory, as well as an exploration of the nature of work at McKinsey & Company. Together, with McKinsey & Company, we conducted a cognitive task analysis to understand the problem-solving domain [7]. Using this analysis, we developed a problem-solving framework representing seven major constructs (e.g. situational awareness, metacognition, decision-making). We examined on-the-job activities at McKinsey & Company to ensure that the structure of our problem-solving framework was aligned with the practical skills and abilities employees engage in at the firm. This work laid the groundwork for scenario development within our simulation.

Third, our problem-solving assessments focus on the process in which users solve and engage in during the task. We do not just look for correct or incorrect answers; instead, we aim to understand how a user solved a problem and what strategies they engaged in to do so. This novel approach to cognitive testing in the hiring domain provides an abundance of information to better assess which candidates are likely to succeed at the company.

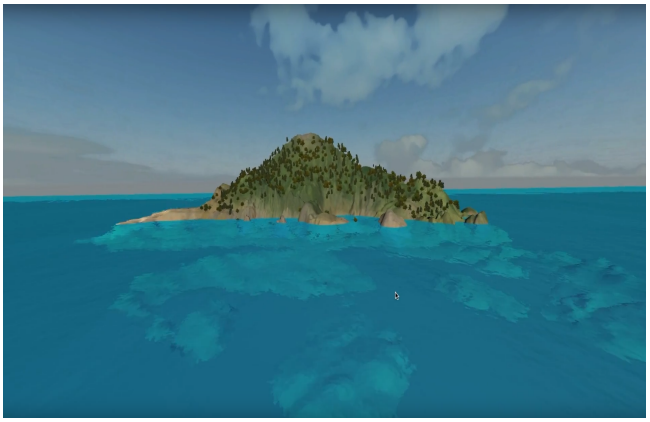

Figure 1. View of natural world simulation environment

We designed each scenario in the assessment based on a set of problem-solving constructs and workplace activities wrapped in a natural world setting. For example, in one scenario, users may be researching and evaluating an infected species in desert terrain. As users play through a scenario, we test them on both their cognitive process and product by capturing their telemetry data. These hovers and clicks are captured as evidence to make inferences about their cognitive processing.

2. OVERVIEW OF SCORE DEVELOPMENT

Imbellus scores were developed using our problem-solving ontology, comprised of approximately 100 constructs, and the cognitive task analysis we conducted with McKinsey & Company. Imbellus scores quantify how users' actions, timestamps, and performance relate to the cognitive constructs within our problem-solving ontology. We derive all Imbellus scores from the users' telemetry data. We then map the scores to one or more problem-solving constructs within our framework.

To create the Imbellus scores, we engaged in a step-by-step process to build, test, and refine each score and its link to the theoretical framework. First, we built expert models for each scenario within our simulation. Expert models help us understand how applicants' cognitive skills manifest in telemetry data. Within our expert models, we outlined the evidence we expected to see in users' behaviors (e.g. efficiency, systematicity) as they complete tasks. We used these evidence statements to develop our Imbellus scores. Following our initial score design, we conducted a series of think-aloud tests aimed at linking specific thinking patterns and behaviors to our scores. We incorporated information from these think-aloud sessions to revise our expert models and scores. We used the initial set of Imbellus scores as a basis for our November 2017 pilot study.

3. PRELIMINARY PILOT OVERVIEW

Using our preliminary Imbellus scores, we conducted a large-scale pilot study in the Fall of 2017. This pilot study tested the predictive capacity of our scores, as well as assessment and simulation environment. We mapped each Imbellus score to one or more of five high level cognitive constructs: critical thinking, decision-making, metacognition, situational awareness and systems thinking. This mapping allows us to build cognitive profiles while also examining the predictive bearing of each score. The pilot study data will be used to inform future designs, validate methodologies, and refine scores.

3.1 Method

Our pilot test, comprised of 527 McKinsey & Company candidates, represented our largest cohort to date. Testing occurred in London, UK from November 13, 2017 through November 17, 2017 and was an optional part of the candidates' interview process with McKinsey & Company. After the conclusion of our game-based assessment, participants completed a survey designed to collect demographic information and user feedback.

Based on survey data, 40% of participants were female, 59% were male, and 1% chose not to provide gender. Based on the Equal Employment Opportunity Commission's guidelines, the ethnic breakdown of the sample was as follows: 52.6% White, 29.7% Asian, 3.9% Hispanic, 4.1% Mixed, 3.3% Black, 2.8% Other, and 3.5% did not specify [2]. Participants' educational backgrounds ranged from humanities-based disciplines to business and engineering. On English proficiency, 56% of the sample reported being a native English speaker, 43% reported being a fluent but non-native English speaker, and 1% reported having a "business-level" proficiency of English.

The participants in our pilot population were given the option of completing our digital assessment after completing the McKinsey & Company Problem-Solving Test (PST), a paper-based assessment. McKinsey & Company administers the PST at proctored test sites. The PST is a traditional cognitive assessment designed to provide insight into applicants' cognitive skills. For the sample of participants who also completed the Imbellus assessment, the proctors told the participants that chose to complete the Imbellus assessment that the outcomes of the assessment would not affect their recruitment process.

Candidates were allotted 60 minutes, the recommended amount of time excluding cases of learner accommodation, to complete the three scenarios. The digital assessment was administered using McKinsey-owned laptop computers in a controlled environment. Along with assessment telemetry and survey data, we collected all scratch paper used by candidates. The assessments took place over the course of 5 days of testing and 29 sessions, none of which experienced significant technical difficulties.

3.2 Creating Construct Profiles

To better understand how participants performed in our assessment, we created a cognitive profile for each participant based on five cognitive constructs: critical thinking, decision-making, metacognition, situational awareness and systems thinking. We already had created theoretical construct affinities for each item score. However, not every item score was predictive. We created a non-negative logistic regression with LASSO regularization to predict the probability a user would pass the first cognitive screen [10].

Before we performed the regression, we imputed missing scores by their median value. All scores were scaled from 0 to 1 using their smallest and largest values. The regression must have non-negative weights because we assume that a higher item score is evidence of higher ability. We used LASSO because of its feature selection properties [11]. The LASSO regularization strength, λ, was found through 10-fold cross-validation. The goal of this step was feature selection, so we chose λ based on a combination of non-zero coefficients and deviance. A λ of 7.68×10^{-3} produced a model with 26 (from 81) non-zero coefficients and a deviance of 1.24 (minimal deviance model = 1.22).

We scaled the resulting item score weights according to their theoretical relevance to each construct. The most relevant scores were multiplied by 3, while relevant scores were multiplied by 2. Marginally relevant scores were not scaled. Item scores that were irrelevant to the construct were set to 0. This created five construct-scaling vectors. The scores for each user were multiplied element-wise by each of the scaling vectors.

These scaled item scores were summed together for each construct. The result was then rescaled by dividing each construct by its highest possible score and transformed into percentile ranks. All construct scores except decision-making had high Pearson correlation (>0.60) with passing McKinsey's multiple-choice Problem-Solving Test (PST). Decision-making had a Pearson correlation of 0.43. The full correlation table between the constructs and passing the PST is displayed below.

Table 1. Correlations between construct scores and PST passing scores

	Meta	**ST**	**SA**	**DM**	**CT**	**PST Pass**
Meta	1.00	0.46	0.52	0.43	0.69	0.63**
ST	0.46	1.00	0.70	0.28	0.80	0.67**
SA	0.52	0.70	1.00	0.28	0.79	0.71**
DM	0.43	0.28	0.28	1.00	0.33	0.43**
CT	0.69	0.80	0.79	0.33	1.00	0.65**
PST Pass	0.63	0.67	0.71	0.43	0.65	1.00

All constructs are significantly related to PST pass rate at p<.01

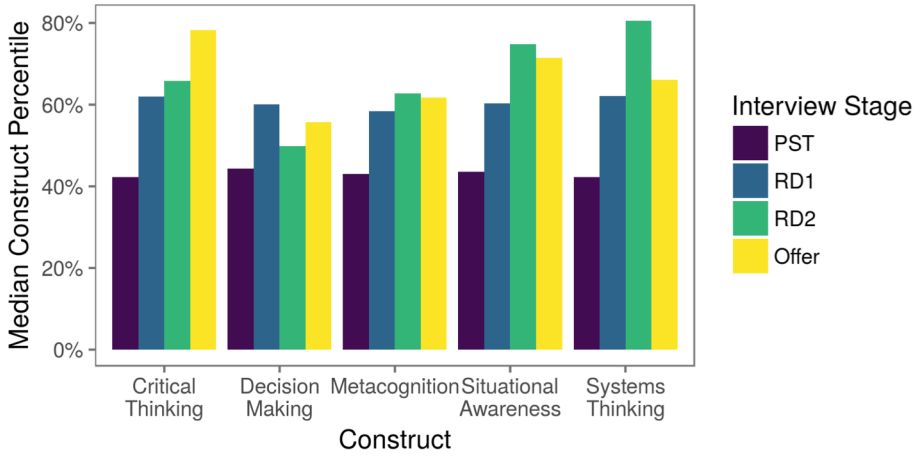

Figure 2. Median Construct Percentile through McKinsey & Company Recruiting Pipeline

The plot above shows the median percentile rank of each of the five construct measures at each stage of the interview process (See Figure 2). Each colored bar in the plot represents the outcome of the interview process. The disposition labeled "PST" signifies that the candidate was screened out before the first interview. "RD1" and "RD2" signify that the applicant did not continue past the first or second round interviews, respectively. "Offer" means that the applicant received an offer from the company.

Below is a table of the median percentile of each of the five constructs at each stage of the interview process along with the median absolute deviation (MAD). This table reveals that preliminary cognitive construct scores are significantly related to success in the interview process. While more work needs to be done to explore this relationship, the initial results are favorable.

Table 2. Median percentile construct score by interview stage.

	PST	RD1	RD2	Offer
Critical Thinking	0.43 (.34)	0.62 (.35)	0.65 (.28)	0.78 (.31)
Decision Making	0.45 (.37)	0.59 (.35)	0.51 (.40)	0.56 (.24)
Metacognition	0.44 (.36)	0.59 (.36)	0.61 (.27)	0.62 (.33)
Situational Awareness	0.44 (.34)	0.6 (.36)	0.74 (.24)	0.71 (.36)
Systems Thinking	0.43 (.35)	0.62 (.35)	0.78 (.28)	0.66 (.19)

Median scores and (Median absolute deviations)

4. CONCLUSIONS & FUTURE WORK

Results from the pilot are promising and show that the Imbellus scores can be used to build out predictive cognitive profiles of candidates. Indeed, these results showed that the cognitive profiles of users were predictive of their success through the McKinsey & Company hiring pipeline. Beyond predictability, these results also show that cognitive processing skills can be captured and quantified using telemetry data within a complex problem-solving task.

To examine the generalizability of these results, we are currently conducting playtests with McKinsey & Company employees and candidates, globally. This extra testing will be used to help us iterate on the design of the assessment and refine our Imbellus scores. In the fall of 2018, we will run a large-scale field test with an expected sample size of over 1000 of McKinsey & Company candidates.

The current version of the simulation is deployed in a secure, proctored environment. In the future, our assessments will be deployed remotely. As such, our assessment will aim to account for performance effects across demographic factors. At its core, Imbellus will leverage a data-driven, artificial intelligence (AI) architecture to prevent cheating. Every user who takes the Imbellus assessment will receive a unique task instance that, on the surface, is varied by its individual properties, complexity, and visual design, while structurally every task version remains consistent in its assessment. Through this approach, Imbellus assessments will prove robust against cheating, hacking, and gaming challenges that face many existing intelligence tests. Our assessments are designed for scale, enabling our team to reach a variety of domains and populations.

Looking beyond this work, we are exploring capabilities beyond problem-solving, including affective skills that are essential for success in the 21st Century workforce. At Imbellus, we aim to provide insightful data points on incoming applicants and current employees that will help companies build successful and sustainable teams in the future.

5. REFERENCES

[1] Assessment & Selection. OPM.GOV. Retrieved from: https://www.opm.gov/policy-data-oversight/assessment-and-selection/other-assessment-methods/cognitive-ability-tests/

[2] Code of Federal Regulations Title 29 – Labor. 1980. Retrieved from: https://www.gpo.gov/fdsys/pkg/CFR-2016-title29-vol4/xml/CFR-2016-title29-vol4-part1606.xml

[3] Gallagher, C. 2003. Reconciling a Tradition of Testing with a New Learning Paradigm. *Educational Psychology Review*. 15, 1, 83-99. DOI=http://www.jstor.org.stanford.idm.oclc.org/stable/23361535

[4] Paul R. Sackett, Nathan R. Kuncel, Justin J. Arneson, Sara R. Cooper, & Shonna D. Waters. 2009. Socioeconomic Status

and the Relationship Between the SAT and Freshman GPA: An Analysis of Data from 41 Colleges and Universities. The College Board, New York. Retrieved from: https://research.collegeboard.org/sites/default/files/publications/2012/9/researchreport-2009-1-socioeconomic-status-sat-freshman-gpa-analysis-data.pdf

[5] Perkins, D. N., & Salomon, G. 1992. Transfer of learning. *International encyclopedia of education*. *2*, 6452-6457.

[6] SAT vs. ACT. The Princeton Review. 2018. Retrieved from: https://www.princetonreview.com/college/sat-act

[7] Schraagen, J. M., Chipman, S. F., & Shalin, V. L. (Eds.). 2000. Cognitive task analysis. *Psychology Press.*

[8] The ACT Test for Students. ACT. 2018. Retrieved from: https://www.act.org/content/act/en/products-and-services/the-act.html

[9] Zanderland, L. 1998. Measuring Minds. *Cambridge University Press*, Cambridge, UK.

[10] Mhukrishnan, R., and R. Rohini. "LASSO: A Feature Selection Technique in Predictive Modeling for Machine Learning." In *2016 IEEE International Conference on Advances in Computer Applications (ICACA)*, 18–20, 2016. https://doi.org/10.1109/ICACA.2016.7887916.

[11] Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, *33*(1), 1–22.