Psychometrically valid game-based assessments

Fact or fiction?

Paul James & Suchi Pathak
Who are we?
- Meet the team..

Why gamify?
- Why games?
- Current state of assessments
- The promise of gaming

How to construct a Game-Based Assessment
- Some possible approaches
- Challenges & considerations
Who are we?

- 35 years
- Business, Education & Sport
- Innovation Team
  - Product development
  - Product launches
  - Bespoke projects
  - Futurology
Paul James

- Head of development
  - ...for a global psychometric test publisher
- Career developer
  - ...with several multi-million projects under my belt
- Innovator
  - Key member of R&D team
  - Conducted research projects with Universities
- Avid gamer
  - ...and amateur game developer

Suchi Pathak

- Head of psychology
  - ...for a global psychometric test publisher
- 10 years in Test Publishing
  - Intellectual property, International markets, Corporate, Military, Sport & Education
- Research and Development
  - Design, validation, adaptation and application, future proofing assessments, university collaborations
- Not an avid gamer!
  - UX Interests - modalities, gamification, big data analytics, machine learning, adaptivity
What is a game?

“A form of competitive activity or sport played according to rules”
- Oxforddictionaries.com

“An activity or contest that has rules and that people do for pleasure”
- Merriam-Webster.com

“A game that is played on a computer”
- Dictionary.cambridge.org

An objective orientated interactive medium based on rules
- Me

...What is a videogame?
Who plays games?
Who plays games?
So... the BIG question
Anatomy of assessments

Psychometric validity

Assessment → Scoring → Results
Useful properties of games

- They assess the gamer
- They generate scores
- They give results at the end
- They are repeatable
- They are based on rules
- ...just add psychometrics
Why gamify? - Assessments

Some drawbacks of questionnaires:

- Impression management
  - Are they telling the truth and/or being consistent?

- Do people like doing assessments?
  - Clicking on a Likert scale

- Constrained by convention
  - Need to be accessible to all

- Abstract
  - Not always directly measuring the subject
  - Based on the rules of questionnaires

Risk Approach

Those who score high on the risk approach tend to be more confident and rational in uncertain situations and consider risk approach characteristics as indicating more instinctual emotional decisions.

How many data points are we capturing?

- Can we capture more?
How else can we assess?
Example

Question: "Would you put your needs before others?"

- for financial gain
- to save a loved one
- to avoid embarrassment
- if your life depended on it
- if everyone knew what you'd done?
- ...would it depend on who the 'others' were?
- ...would you feel guilty afterwards?
- Etc, etc

Wrex: This base can't be destroyed. I won't allow it!
These aren't your people!
[Signal Ashley to kill Wrex.]
Don't be an idiot!
Calm down.
We can work this out.
[Shoot Wrex.]
These aren't your people!

[Signal Ashley to kill Wrex.]

Calm down.

We can work this out.

[Shoot Wrex.]

Don't be so naive.
What are your thoughts on this game as a measure of empathy?
Create a relevant scenario for the assessment in hand
  - Help them relate to the question you’re asking
  - Assess directly by observing behaviour

Provide a consistent narrative
  - Makes the assessment more palatable
  - Helps the player identify with the context of the questions

Help the player get in to the flow
  - Recording instinctive reactions
  - Helping them identify with their avatar
Multiple metrics

Games are great platforms for recording data

- They already model everything the player does
- Validating the game itself
- Validating the player’s claims
- GBAs have many more data points than conventional assessments. Big data has been proven to predict personality and many other outcomes.

Machine Learning Algorithms

8 million Facebook users

Films

Music

Food
Good games are engaging
- Billions of people choose to play games

Positive experience
- Candidates actually enjoying being assessed!
- Higher completion rates?

Positive after impression
- Brand awareness and loyalty
- Viral sharing
- Easier to find willing volunteers
How to construct a game based assessment
Principles of GBA design

- **Blended approach**
  - Use elements of valid psychometrics and gaming to construct your assessments

- **‘Big data’ philosophy**
  - Develop as many different metrics as you can
  - Draw conclusions from multiple sources
  - No bad ideas. Test everything and iterate

- **Equivalence chaining**
  - Use a robust starting point as a yardstick & keep validating

- **Make good games**
  - Must be engaging
  - Must be technically competent
  - Know your target market

- **Play more games…**
Process flow for creating a GBA

Psychologists

Planning & Research

Create psychometric questionnaire

Game developers

Games industry research

Game design principles

Psychometric design principles

Blue sky / conversion

Prototype

Both

Validation

GBA!
Process flow for creating a GBA

Planning & Research

Games industry research

Game design principles

Psychometric design principles

Blue sky / conversion

Prototype

Validation

GBA!
Things to be wary of

- **Brand new field**
  - You will be breaking new ground

- **Budget and scope**
  - Games are more expensive to produce
  - Gamifying is not always appropriate

- **Gaming the system**
  - Be wary of exploits and FOO strategies
  - Make sure that the assessment is not the main objective of the game
    - There must be no direct or tangible reward for choosing one action over another

- **Inbuilt skill bias**
  - Does previous knowledge distort scores?
What are the dangers of GBAs?

Ethical challenges

GBA challenges are similar to Big Data challenges

- Will people know they are being assessed?
- Danger of linking data together without a clear objective/hypothesis

Are the questions being framed correctly?

- Are people getting pigeon-holed into choosing a set of responses?
- Does this result in contaminated data or negative candidate perceptions?
- Are there too many variables with GBAs?

Inclusivity

- Is the perception of gamers or accessibility of games disadvantaging some demographic groups?
- Do people want to play games?
Thank You!

Paul James
Paul@thomasdev.co.uk

Suchi Pathak
Suchip@thomas.co.uk
Psychometrically valid game-based assessment

Fact or Fiction?
Introducing...

■ Lara Montefiori
  - Head of Psychology at Arctic Shores
  - PhD student at UCL
  - Main interests:
    - Disrupt and Innovate the assessment market
    - Find an alternative to self-report
    - Have fun doing the above

■ Arctic Shores
  - Est. 2014
  - Growing fast
  - Team diversity
  - Clients portfolio
What do we do?

- Actions speak louder than words...

- LET’S PLAY!
Kevin (35, male)
- Alpha male syndrome
- Always starts with ROCK because it’s STRONG

Mylo (8, male)
- Very creative kid
- Always chooses WELL because it’s UNIQUE (secret German version)

Alba (5, female)
- Not quite the ultimate strategist, yet
- Always repeat the move that won before

Lara (36, psychologist)
- Single shot, choose paper with Kevin and Mylo
- Best of three with Alba (50:50, win, win)
- I ALWAYS WIN!
What do we really do?

- Firefly Freedom
  - First game
  - Five Levels
  - Common aim to catch Fireflies
  - Developed as a measure of INNOVATION POTENTIAL
Firefly Freedom
Yellow Hook Reef

- Pirate game, three levels so far
- About 40 minutes
- Measures Abstract and Verbal Reasoning, and Numeracy
Captain, why can't we store fresh food and drink on the ship?

Ships can't maintain the cool temperatures needed

Ships are too open to allow fresh food on-board

Ships lack the required space to store fresh food

Ships would smell if fresh food was kept on-board

Food and Drink Inventory

Food and drink are of vital importance when travelling aboard a ship. If the supply of food and drink runs out this would damage crew morale and the running of the ship, causing a big problem for any captain.

Fresh vegetables, fruit and meat are unsuitable on long voyages as, due to the lack of refrigeration,
Yellow Hook Reef – Abstract Reasoning
Yellow Hook Reef - Numeracy

0:43

4 16 64 128 256

7 8 9
4 5 6
1 2 3
0  ENTER
 Cosmic Cadet

- Collection of mini games
- Set in Space
- About 20 minutes long
- Measures 13 traits, clustered into 4 factors

<table>
<thead>
<tr>
<th>Information Processing</th>
<th>Delivering Results</th>
<th>People Orientation</th>
<th>Thinking Style</th>
</tr>
</thead>
<tbody>
<tr>
<td>Executive Function</td>
<td>Resilience</td>
<td>Social Confidence</td>
<td>Learning Agility</td>
</tr>
<tr>
<td>Processing Speed</td>
<td>Persistence</td>
<td>Affiliativeness</td>
<td>Risk Appetite</td>
</tr>
<tr>
<td>Attention Control</td>
<td>Performance Under Pressure</td>
<td></td>
<td>Managing Uncertainty</td>
</tr>
<tr>
<td>Processing Capacity</td>
<td></td>
<td></td>
<td>Innovation Potential</td>
</tr>
</tbody>
</table>
Cosmic Cadet
Skyrise City

- More grounded graphics
- Less storyline
- Different model altogether
- Merging personality, cognition, ability, and potential.
- Pick and mix from
  - Six factors
  - Sixteen Dimensions
  - Fifty-nine Constructs
- Mapped on clients’ Competency Frameworks
- Select relevant levels
- Generates FIT SCORE
GBA’s Principles

- Observation of Behaviour in controlled context
GBA’s Principles

- Observation of Behaviour in controlled context
Para-data

Please read and agree to our terms and conditions to continue. (Drag to scroll)

APPLICATION TERMS AND CONDITIONS

PLEASE READ CAREFULLY BEFORE ACCESSING THE APPLICATION

This Application (defined below) is published by or on behalf of Arctic Shores Limited (Company No. 8589048) whose registered office is at 17 Marble Street, Manchester M2 3AW, United Kingdom ("Arctic Shores", "we" or "us"). Our VAT number is 165182604.

Disagree

Agree
- **Average game session about 3000 data points**

- **Record to date 13000+**

- **Enables**
  - Corroboration of scores
  - Really sophisticated cheat traps
  - Comprehensive predictions
  - New hypotheses

- **Enough for**
  - Machine Learning
  - Artificial Intelligence
What does it mean for a psychometric test to “work”? 

- Reliability 
- Validity 
- Resistance to distortion 
- Adverse impact 
- Candidate’s perceptions 
- Effect on Employer’s image
Difficult to measure Cronbach’s Alpha due to nature of the data

Parallel forms – great but expensive solution

Test-retest better solution

People trying to cheat even better
Construct Validity

- Only way is self-report
- IS self-report valid?

Apples and Pears

- iOS vs Android
- Validation is necessary but not feasible

Predictive Validity is crucial

- Future developments
  - Clients
  - Graduates
Resistance to distortions

- **iOS vs Androids**
  - Game play data was identical

- **Low, medium, and high stakes**
  - As long as there is a reward (as little as a chance to win £10), game play data of applicants and testers does no differ

- **Study on Impression Management**
  - Testers paid to fit a profile
  - Failed in GBA but succeeded in self-reported measure
No adverse impact was found (5000+ sample) in:

- Game experience, Age, Gender, Ethnicity, Screen size, Handedness, and Disability (after working memory adjustment for Dyslexia)

Some differences, but all explained by interactions or actual differences:

- Non-white players less persistent – due to iOS not to ethnicity
- Older players lower Working Memory – cognitive decline
- People with more school titles higher on persistence – not surprising
Candidates’ Perceptions – Employers’ Image

- Large Survey Comparing GBA to Self-reports (2000+ relevant sample)

<table>
<thead>
<tr>
<th>GBAs better than SRs</th>
<th>GBAs and SRs are equal</th>
<th>SRs better than GMAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face validity</td>
<td>46.2%</td>
<td>36.2%</td>
</tr>
<tr>
<td>Fairness Perception</td>
<td>42%</td>
<td>36.5%</td>
</tr>
<tr>
<td>Brand Image</td>
<td>47%</td>
<td>41.8%</td>
</tr>
</tbody>
</table>

- More in detail
  - 60% believe that GBAs are harder to fake than SR
  - 40% believe that GBAs measure performance better than SR, or as good (22%)
  - 62% believe that GBAs are better than SR at capturing aspect of their personality of which they are not aware
  - 80% consider employers using GBAs more innovative
  - 70% consider them equally as professional as those using SRs
Challenges

■ Internal
  ▪ Team diversity
  ▪ Niche of expertise
  ▪ Infrastructure
  ▪ Cost
  ▪ Testing
  ▪ Timelines

■ External
  ▪ Skepticism
  ▪ BPS Test Review
  ▪ Perception
  ▪ Complaints
Summary

- What are GBAs
- How do they innovate the Industry
- Advantages
- Disadvantages
- More research is needed
So... Is Game-Based Assessment FACT or FICTION?
Verdict...
Thank you! And don’t forget to PLAY everyday!