Game Based Assessments

Are they really the future?

12 May, 2019

Prepared by:

STEN10
Ben Williams
Business Psychologist

Kings Head House, 15 London End,
Beaconsfield HP9 2HN

☎ +44 (0)1494 412 861
📱 +44 (0)7939 156 708
✉ ben@sten10.com/amy@sten10.com

Who I am

- Chartered Psychologist
- Managing Director of Sten10 Ltd. / Chair of ABP
- Publisher-independent
- (Was an) avid gamer

sten10
experts in people assessment

Agenda

LEVEL 1 - Introduction to Game Based Assessment

- Key parameters of a GBA
- Four types of GBA

LEVEL 2 - Evidence Base

- Types of Evidence
- Reliability / Validity / Adverse impact / Engagement

LEVEL 3 - Conclusions

sten10
experts in people assessment

Level 1
Introduction to GBA

sten10

experts in people assessment

Key Parameters of a GBA

- Nature: Gamification vs. Game Based Assessment
- Type: Custom-built vs. pre-existing vs. gamified traditional vs. VR
- Measures: performance, behavioural choice and / or ‘meta-data’ to assess:

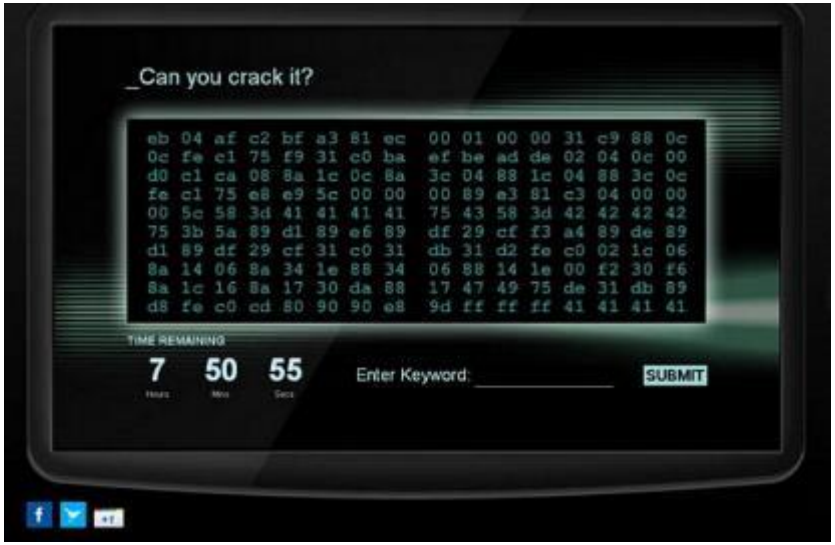
  - Abilities:
    - Cognitive processing speed
    - Attention span
    - Working memory
    - V, N, A reasoning

  - Personality traits:
    - Persistence
    - Risk propensity
    - Emotional Intelligence

  - ‘Role-Fit’ – A.I. % match

sten10
experts in people assessment

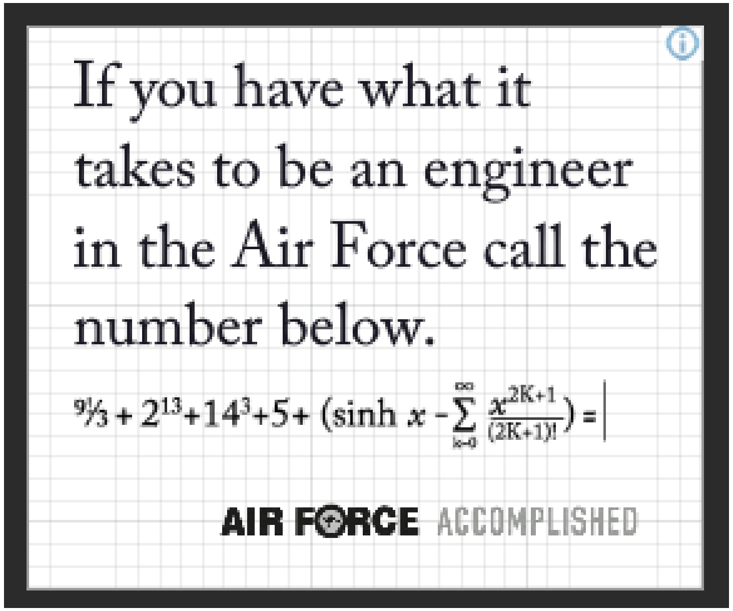
Gamification in Recruitment



_Can you crack it?

```
eb 04 af c2 bf a3 81 ec  00 01 00 00 31 c9 88 0c
0c fe c1 75 f9 31 c0 ba  ef be ad de 02 04 0c 00
d0 c1 ca 08 8a 1c 0c 8a  3c 04 88 1c 04 88 3c 0c
fe c1 75 e8 e9 5c 00 00  00 89 e3 81 c3 04 00 00
00 5c 58 3d 41 41 41 41  75 43 58 3d 42 42 42 42
75 3b 5a 89 d1 89 e6 89  df 29 cf f3 a4 89 de 89
d1 89 df 29 cf 31 c0 31  db 31 d2 fe c0 02 1c 06
8a 14 06 8a 34 1e 88 34  06 88 14 1e 00 f2 30 f6
8a 1c 16 8a 17 30 da 88  17 47 49 75 de 31 db 89
d8 fe c0 cd 80 90 90 e8  9d ff ff ff 41 41 41 41
```

TIME REMAINING
7
Hours
50
Mins
55
Secs
Enter Keyword:
SUBMIT



If you have what it takes to be an engineer in the Air Force call the number below.

$9/3 + 2^{13} + 14^3 + 5 + (\sinh x - \sum_{k=0}^{\infty} \frac{x^{2K+1}}{(2K+1)!}) =$

AIR FORCE ACCOMPLISHED

sten10
experts in people assessment

Types of GBA

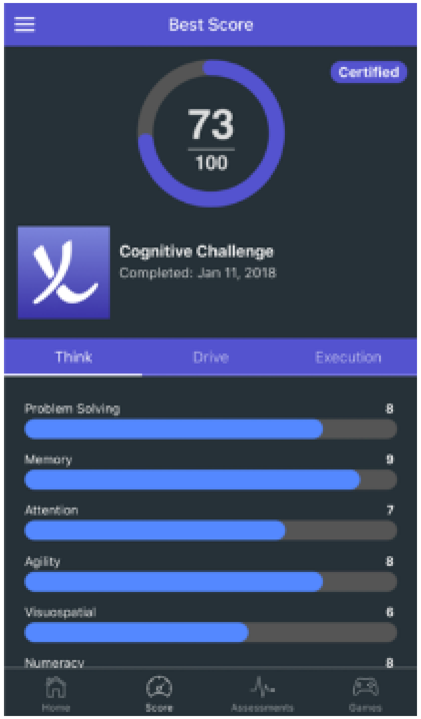
1. Custom-Built GBA's

sten10
experts in people assessment

Arctic Shores

Knack

HireVue (formerly MindX)



Best Score

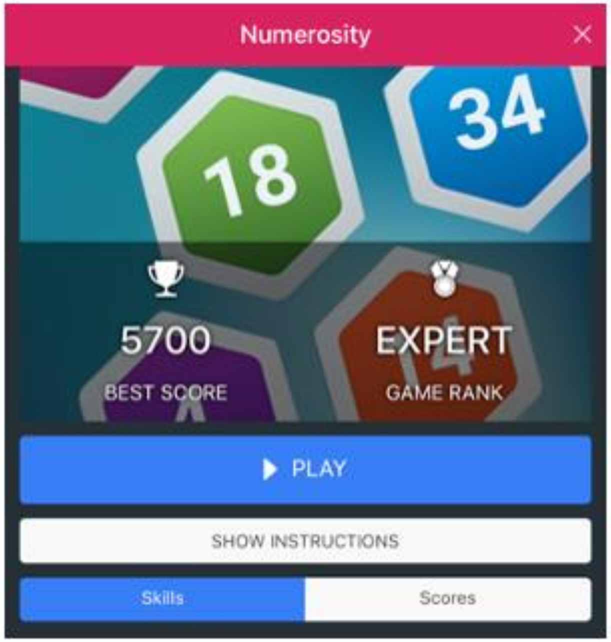
Certified

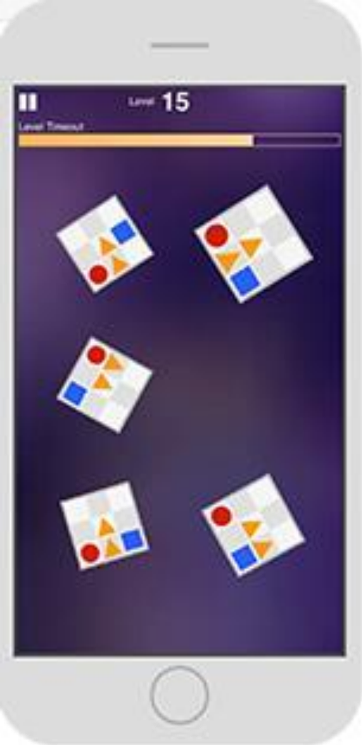
73
100

Cognitive Challenge
Completed: Jan 11, 2018

Think | Drive | Execution

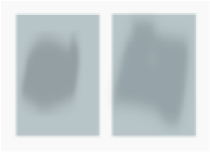
Problem Solving	8
Memory	9
Attention	7
Agility	8
Visuospatial	6
Numeracy	8



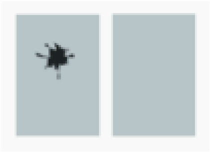


sten10
experts in people assessment

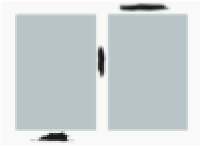
Quest



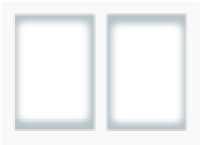
Dirty glass



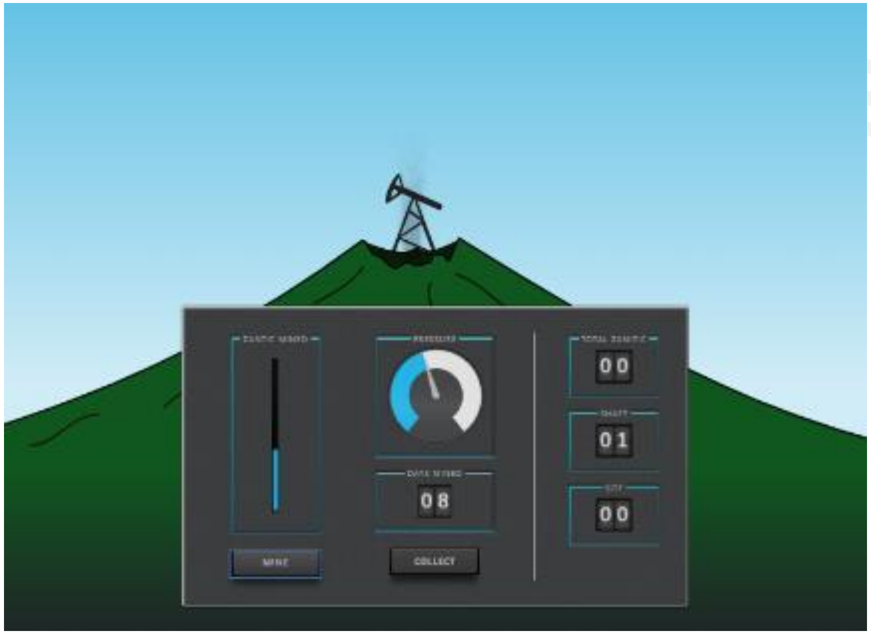
Broken glass

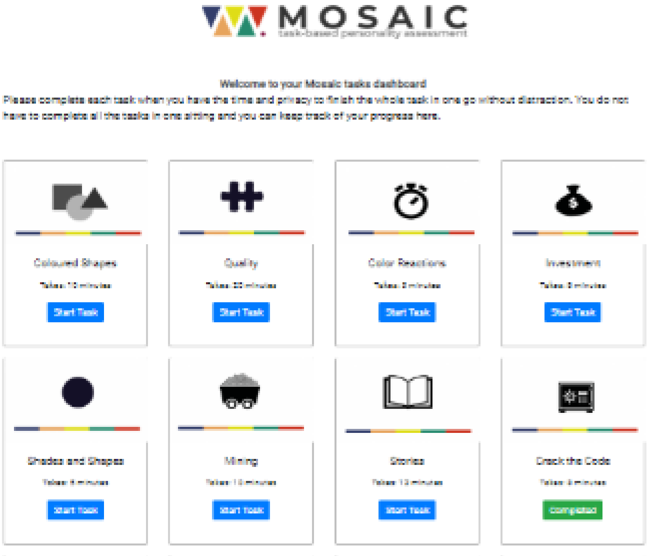


Damaged frame



Faded panes





MOSAIC
task-based personality assessment

Welcome to your Mosaic tasks dashboard
Please complete each task when you have the time and privacy to finish the whole task in one go without distraction. You do not have to complete all the tasks in one sitting and you can keep track of your progress here.

Coloured Shapes
Takes: 10 minutes
Start Task

Quality
Takes: 20 minutes
Start Task

Color Reactions
Takes: 5 minutes
Start Task

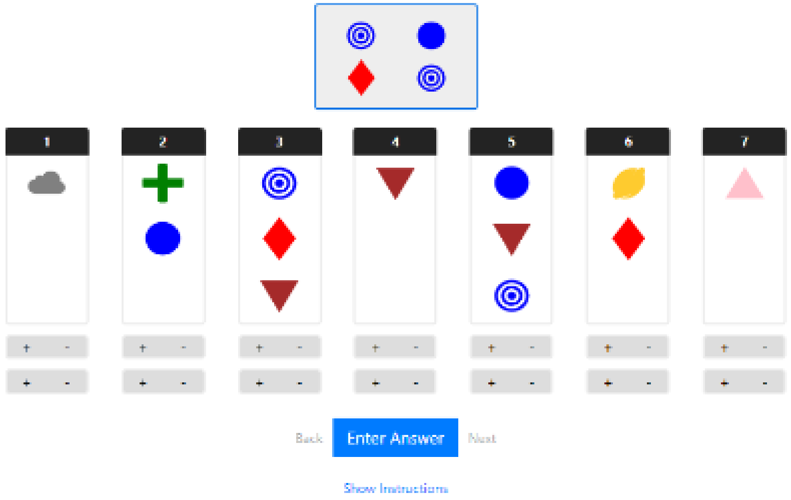
Investment
Takes: 5 minutes
Start Task

Shades and Shapes
Takes: 6 minutes
Start Task

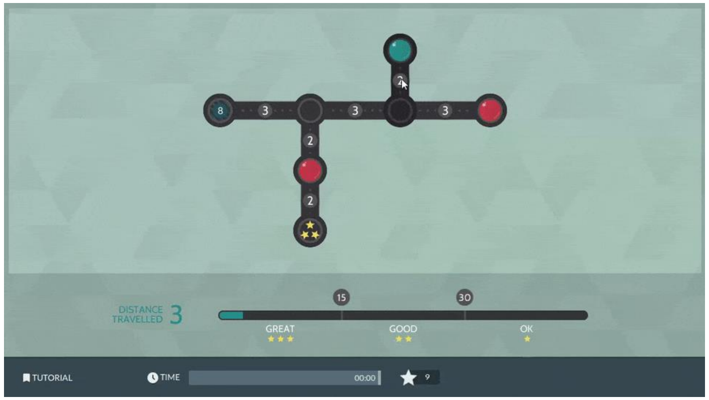
Mining
Takes: 10 minutes
Start Task

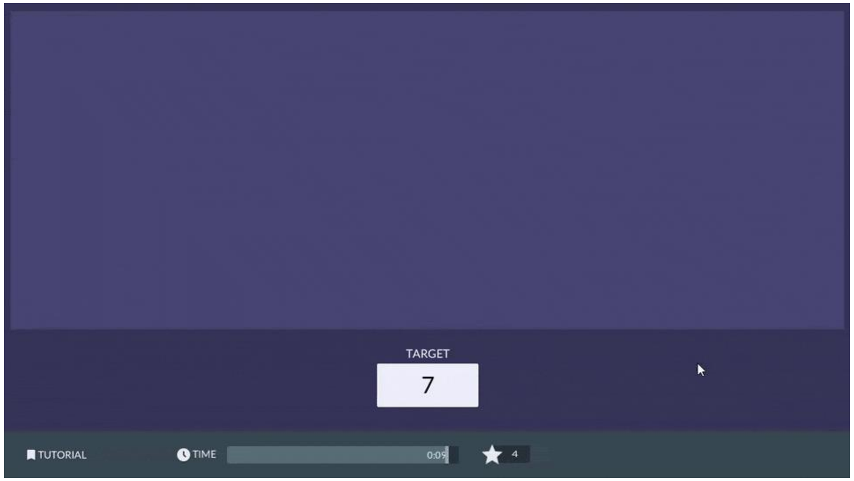
Stories
Takes: 12 minutes
Start Task

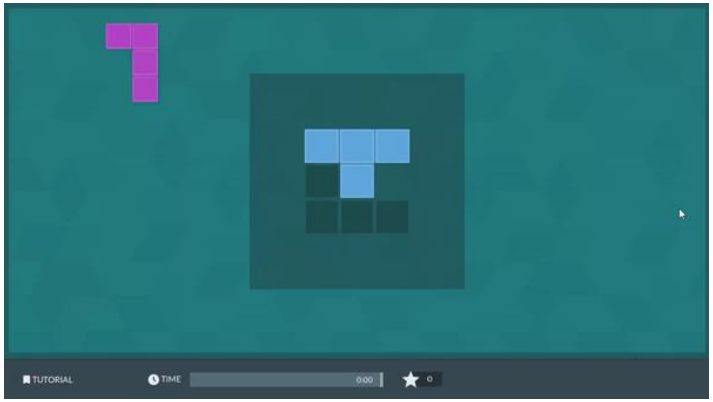
Crack the Code
Takes: 4 minutes
completed

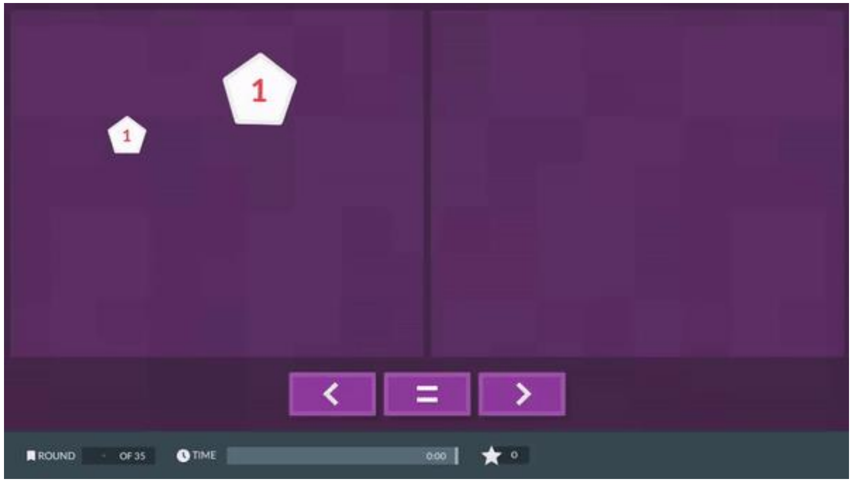


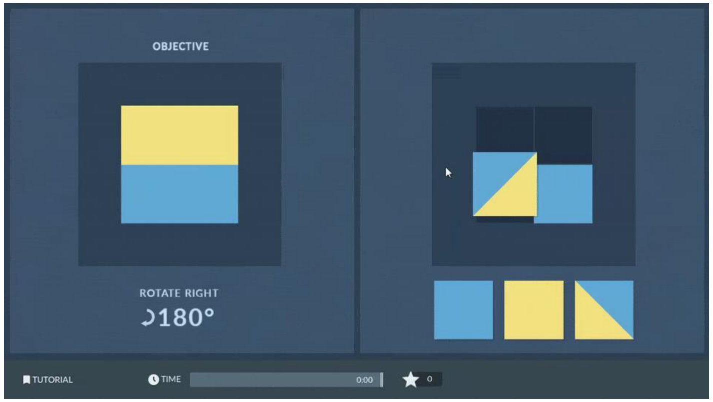
Revelian


8
3
3
3
2
2
DISTANCE TRAVELLED 3
15
30
GREAT
GOOD
OK
TUTORIAL
TIME
00:00
9

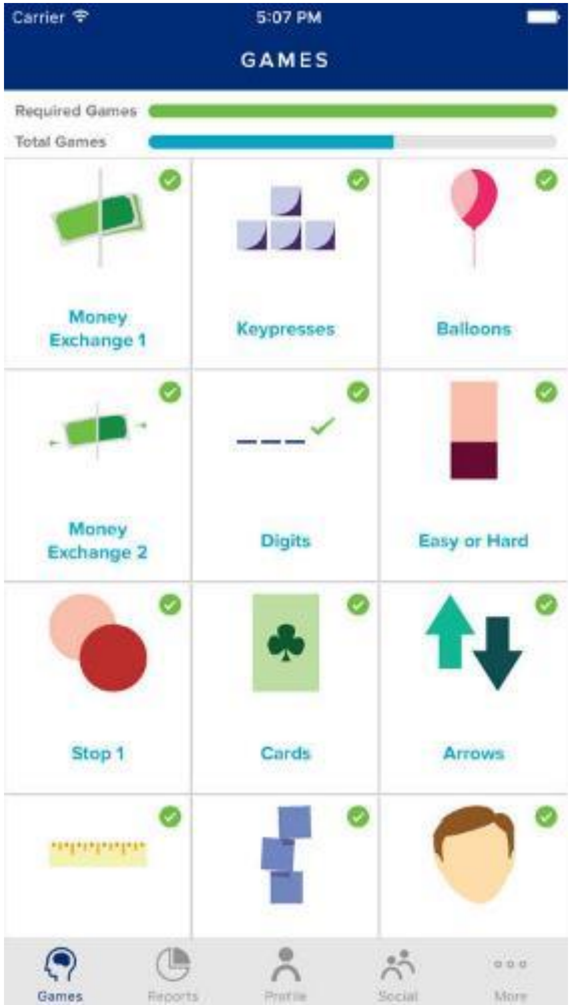

TARGET
7
TUTORIAL
TIME
0:09
4

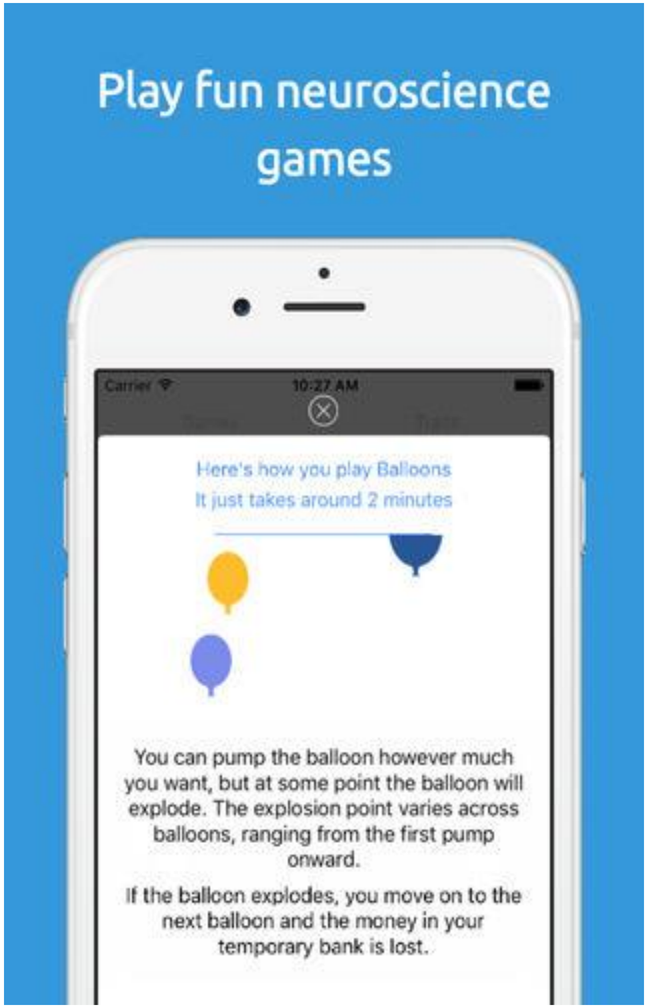

TUTORIAL
TIME
0:00
0


1
1
<
=
>
ROUND
OF 35
TIME
0:00
0


OBJECTIVE
ROTATE RIGHT
180°
TUTORIAL
TIME
0:00
0

Pymetrics

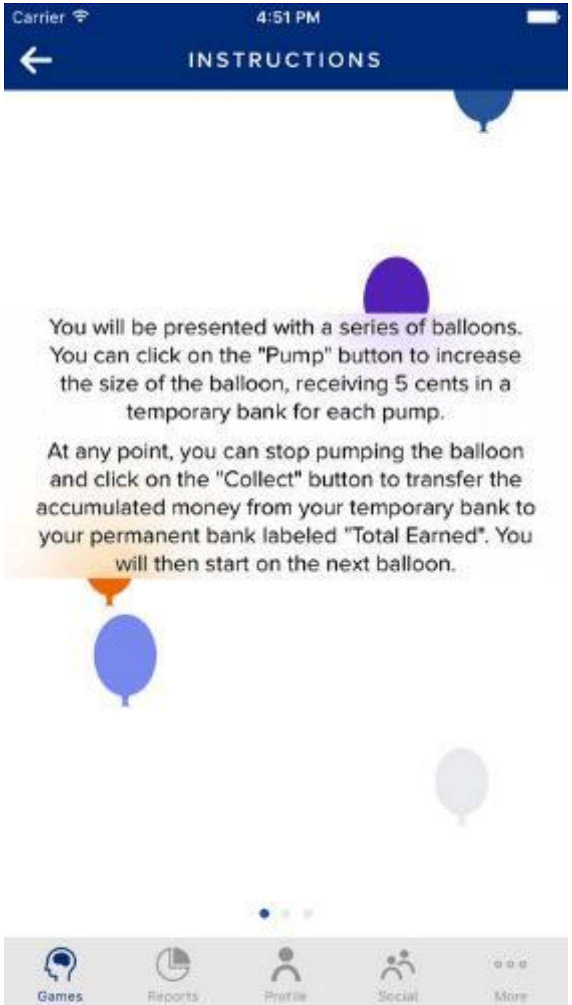



Play fun neuroscience games

Here's how you play Balloons
It just takes around 2 minutes

You can pump the balloon however much you want, but at some point the balloon will explode. The explosion point varies across balloons, ranging from the first pump onward.

If the balloon explodes, you move on to the next balloon and the money in your temporary bank is lost.



You will be presented with a series of balloons. You can click on the "Pump" button to increase the size of the balloon, receiving 5 cents in a temporary bank for each pump.

At any point, you can stop pumping the balloon and click on the "Collect" button to transfer the accumulated money from your temporary bank to your permanent bank labeled "Total Earned". You will then start on the next balloon.

Types of GBA

2. Pre-existing

sten10
experts in people assessment

‘Pre-Existing’ Games

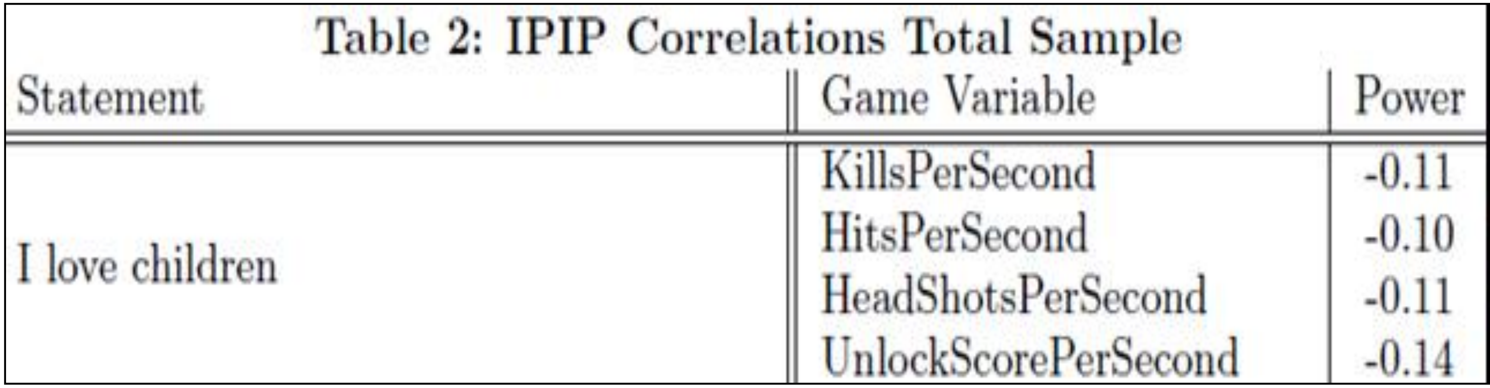


Table 2: IPIP Correlations Total Sample

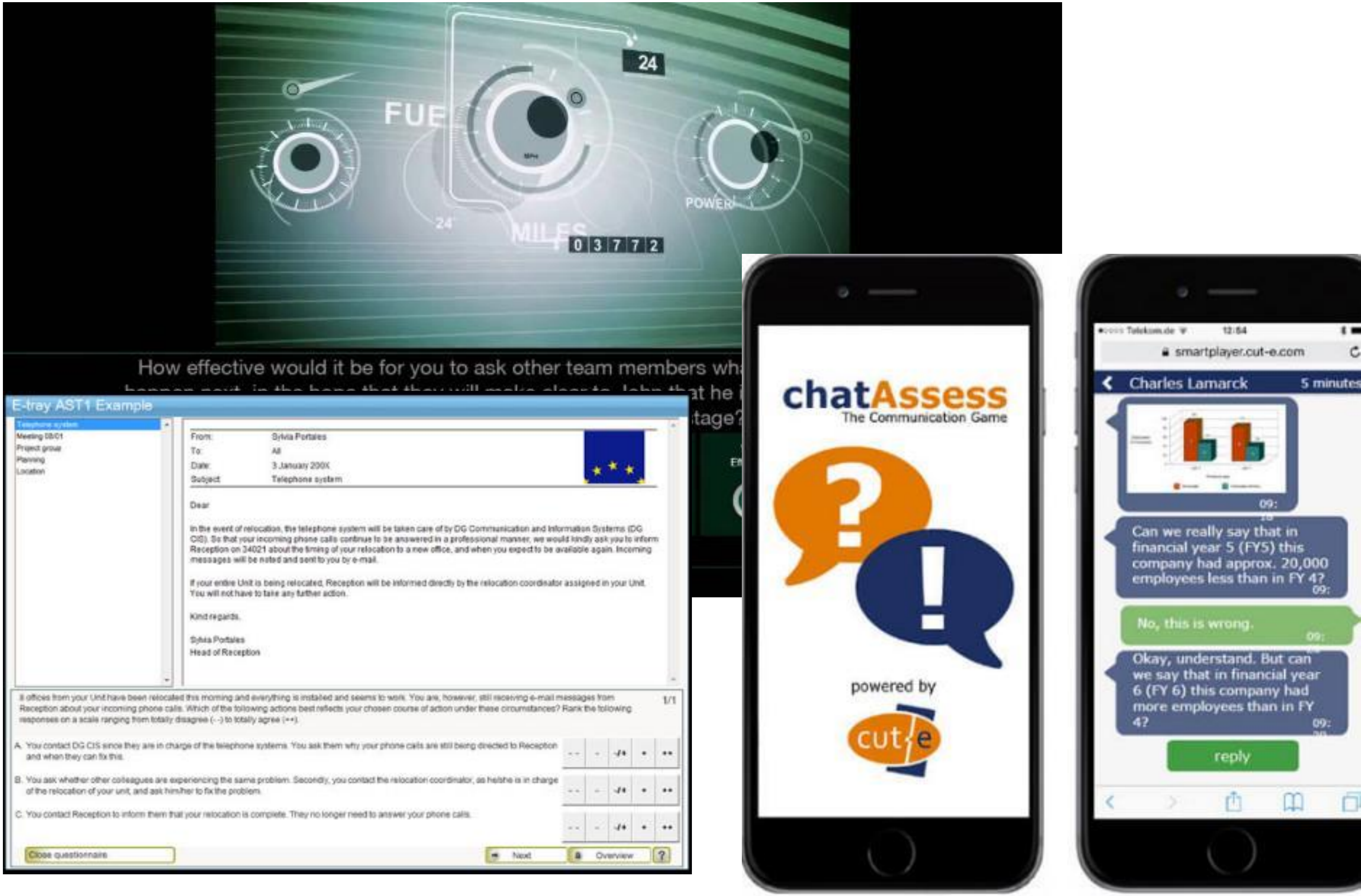
Statement	Game Variable	Power
I love children	KillsPerSecond	-0.11
	HitsPerSecond	-0.10
	HeadShotsPerSecond	-0.11
	UnlockScorePerSecond	-0.14

Types of GBA

3. Tailored Traditional

sten10
experts in people assessment

Gamified Assessments (Not ‘Games’?)



How effective would it be for you to ask other team members wh

happen next, in the hope that they will make clear to John that he i

tage?

E-tray AST1 Example

From: Sylvia Portales
To: All
Date: 3 January 200X
Subject: Telephone system

Dear

In the event of relocation, the telephone system will be taken care of by DG Communication and Information Systems (DG CIS). So that your incoming phone calls continue to be answered in a professional manner, we would kindly ask you to inform Reception on 34021 about the timing of your relocation to a new office, and when you expect to be available again. Incoming messages will be noted and sent to you by e-mail.

If your entire Unit is being relocated, Reception will be informed directly by the relocation coordinator assigned in your Unit. You will not have to take any further action.

Kind regards,

Sylvia Portales
Head of Reception

All offices from your Unit have been relocated this morning and everything is installed and seems to work. You are, however, still receiving e-mail messages from Reception about your incoming phone calls. Which of the following actions best reflects your chosen course of action under these circumstances? Rank the following responses on a scale ranging from totally disagree (- -) to totally agree (++). 1/1

A. You contact DG CIS since they are in charge of the telephone systems. You ask them why your phone calls are still being directed to Reception and when they can fix this. - - | - | -/+ | + | ++

B. You ask whether other colleagues are experiencing the same problem. Secondly, you contact the relocation coordinator, as he/she is in charge of the relocation of your unit, and ask him/her to fix the problem. - - | - | -/+ | + | ++

C. You contact Reception to inform them that your relocation is complete. They no longer need to answer your phone calls. - - | - | -/+ | + | ++

chatAssess
The Communication Game

powered by

cut-e

Charles Lamarck 5 minutes

09:

Can we really say that in financial year 5 (FY5) this company had approx. 20,000 employees less than in FY 4? 09:

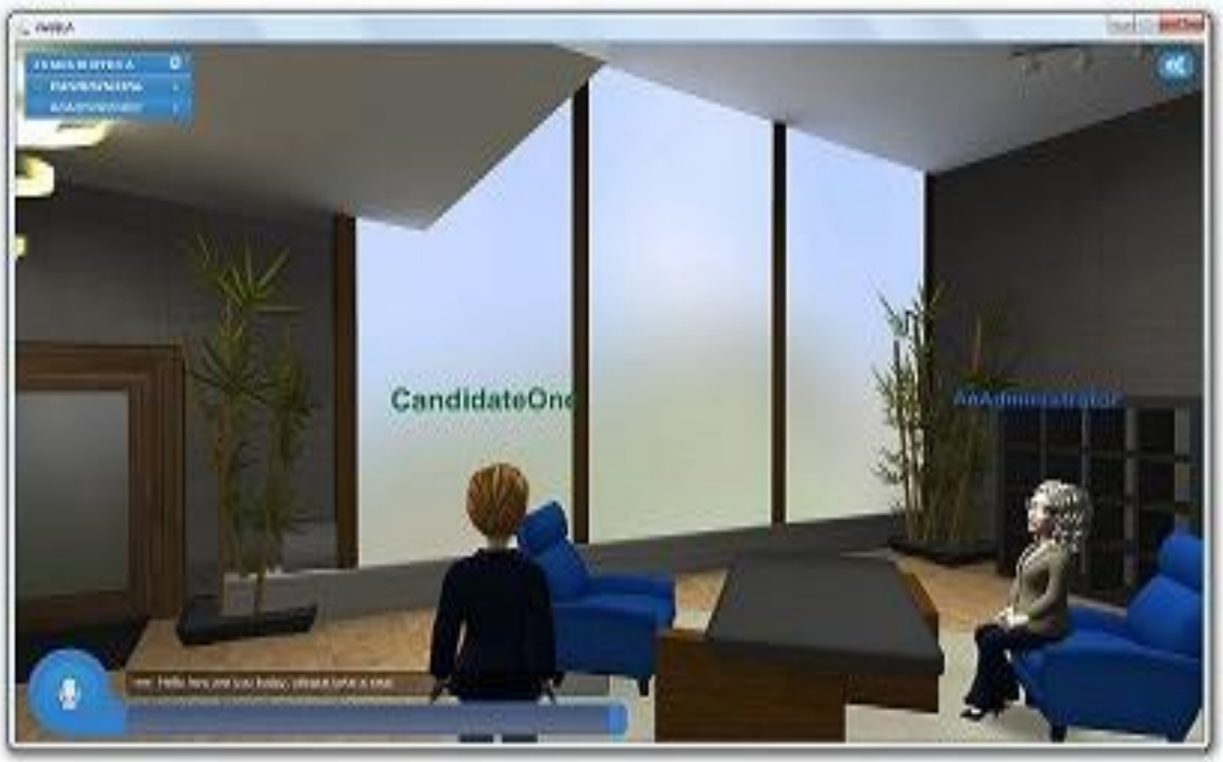
No, this is wrong. 09:

Okay, understand. But can we say that in financial year 6 (FY 6) this company had more employees than in FY 4? 09:

Types of GBA

4. Virtual Worlds, Virtual Reality

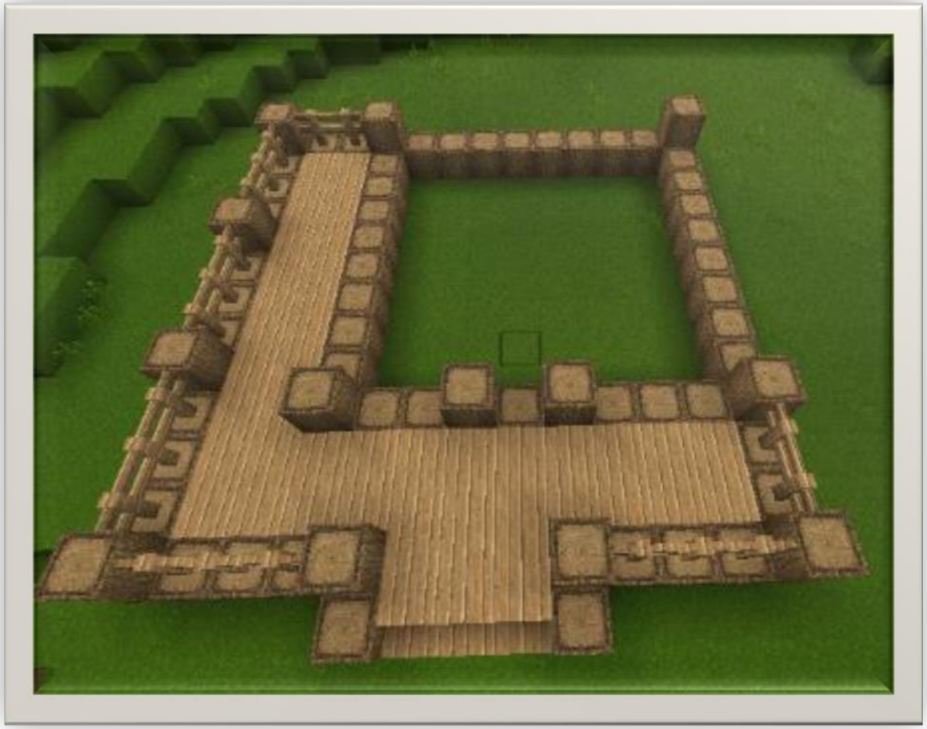
sten10
experts in people assessment



Hello how are you today, please take a seat

sten10
experts in people assessment

Level 2
Evidence
Base



sten10
experts in people assessment

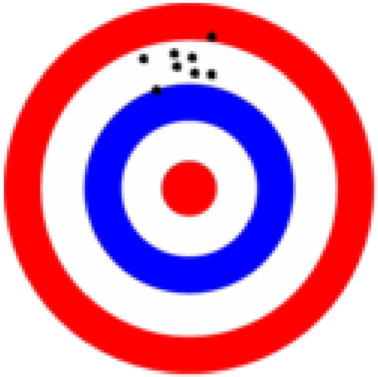
The Challenges

The challenges of establishing psychometric properties:

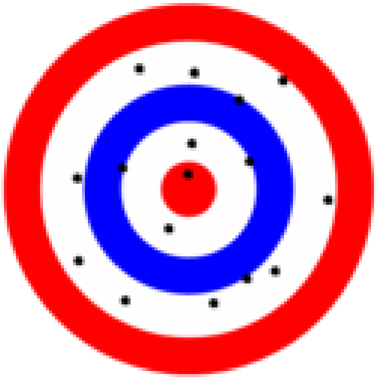
- **A New Market** - GBA Test publishers are quite young meaning evidence of predictive power is limited by necessity
- **Generalisations** about the evidence base are difficult compared to ‘traditional’ psychometrics due to the variety of design
- **Objectivity** - Investigating GBAs objectively is problematic as commercial IP is tied up in the algorithms used. Also, most research being funded and facilitated by the publishers themselves
- **Common method variance** – using GBAs changes the way constructs are measured (construct validity)
- **Complex** – not only raw score but thousands of meta-data points are measured

sten10
experts in people assessment

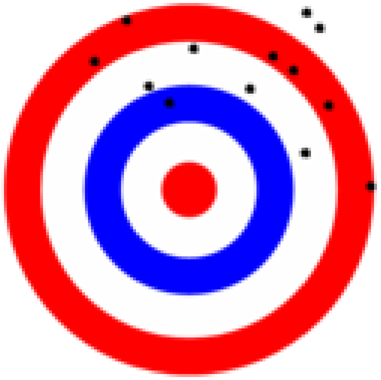
Reliability and Validity



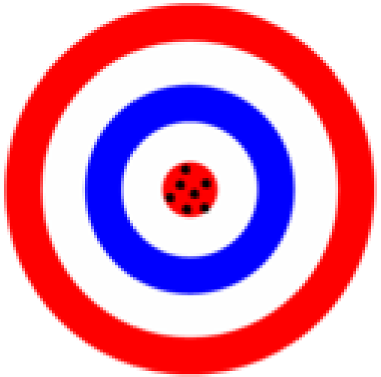
Reliable not Valid
Precise not Accurate



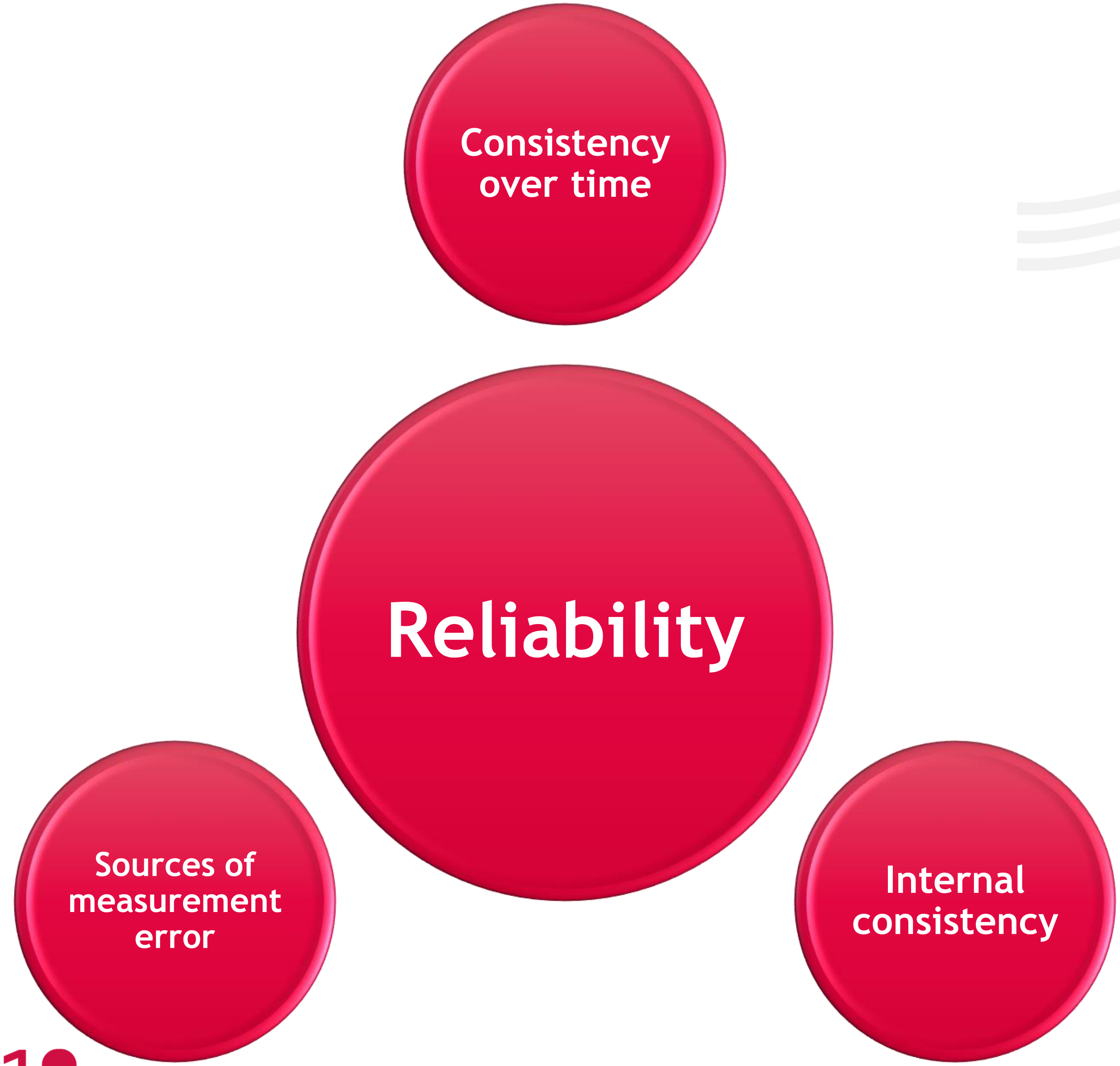
Not Reliable but Valid
Not Precise but Accurate



Not Reliable and not Valid
Not Precise and not Accurate



Reliable and Valid
Precise and Accurate



Consistency over time

Reliability

Sources of measurement error

Internal consistency

experts in people assessment

Consistency

(All from test GBA test publishers)

Internal consistency

- 0.6 – 0.9 (n = 6,000)
- 0.51 – 0.96 (n = < 100)
- 0.84 (n = 500)

(n.b. typical vs maximum ideal values)

Consistency over time

- 0.57 – 0.82 test-retest

Parallel form

- 0.44 – 0.79 for subtests
- >0.9 for app version vs laptop version

sten10
experts in people assessment

Sources of Measurement Error

Length of assessment

- Greater engagement: longer assessment: better reliability? (Riley, 2015)

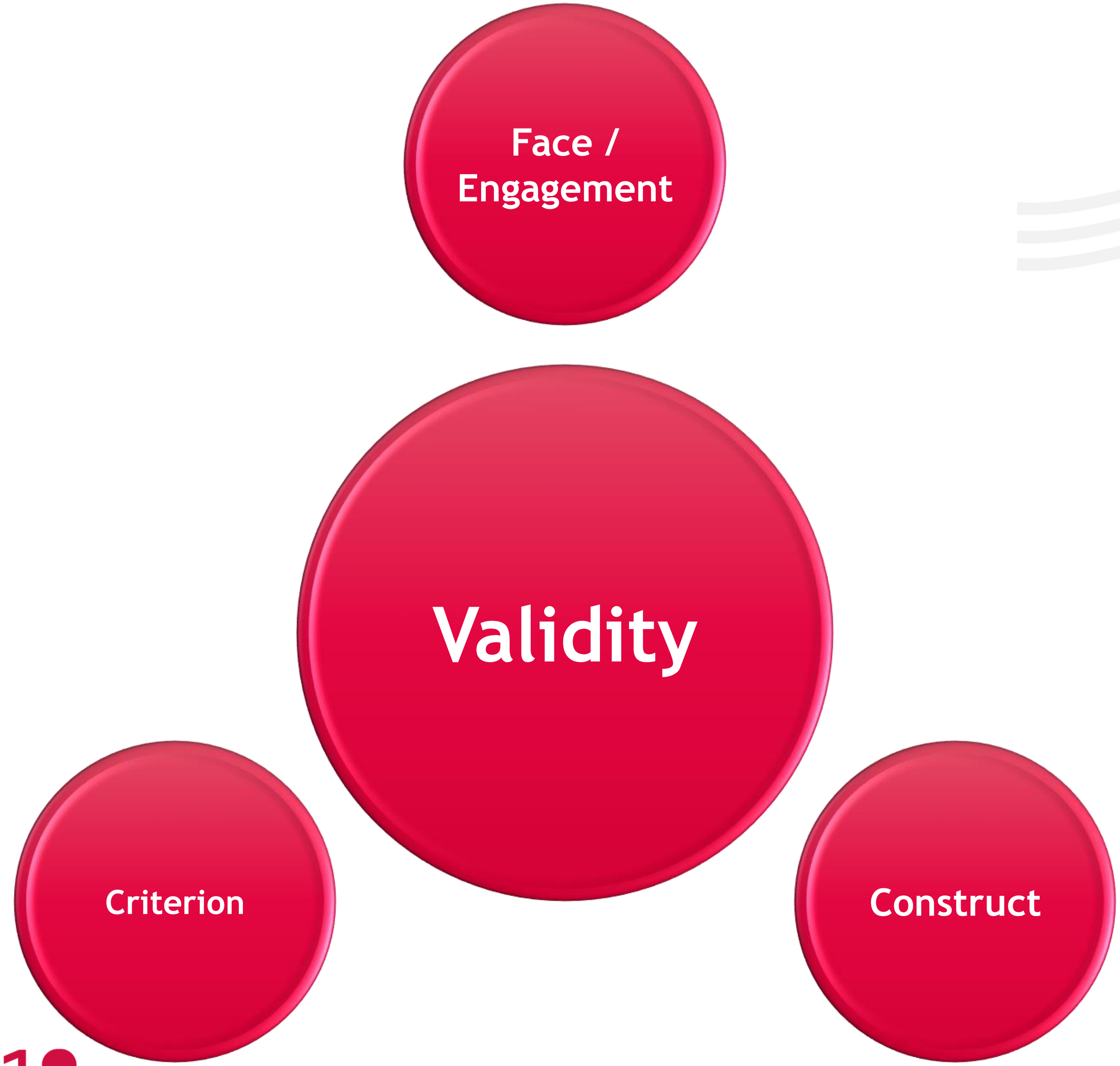
Distortion

- GBA assesses behaviour directly, not through self report: more resistant to distortion? (Landers, 2015) Scores modified on self-report PQs for extraversion and agreeableness, but unable to in a GBA (Montefiori, 2016)

Irrelevant Factors

- Potential reliance on irrelevant factors such as hand-eye co-ordination. Highly interactive games may create unnecessary cognitive load. (Zapata-Rivera & Bauer, 2012)

sten10
experts in people assessment



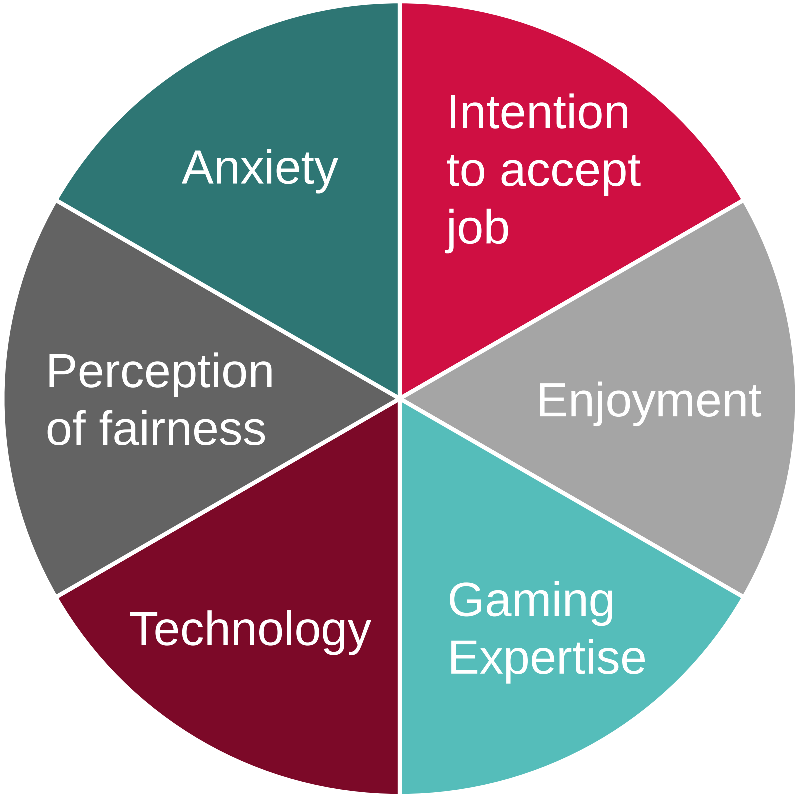
Face / Engagement

Validity

Criterion

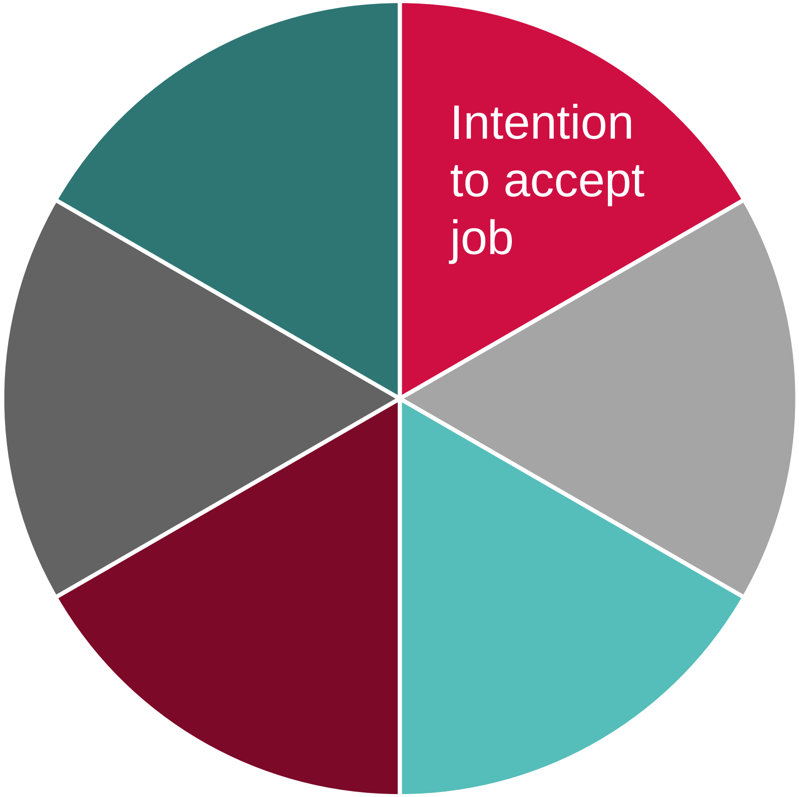
Construct

sten10
experts in people assessment



Face Validity / Engagement

- Selected studies



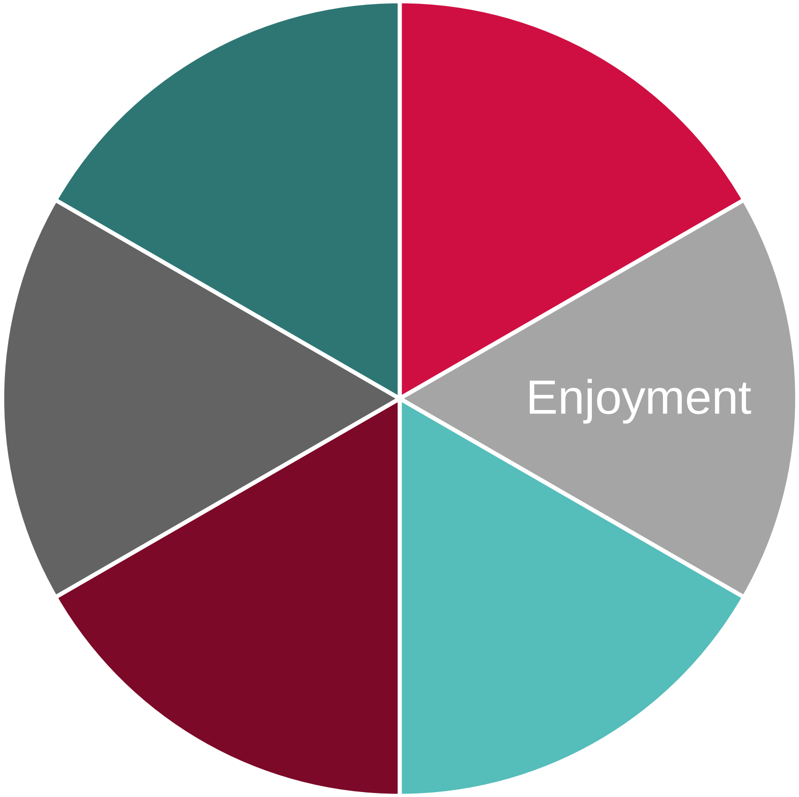
Intention to accept job offer

Animated characters = positive attitude towards hiring company, stronger intention to accept a job offer (e.g. Motowidlo et al., 1990; Richman-Hirsch et al., 2000; Bruk-Lee et al., 2012)

sten10 experts in people assessment

Face Validity / Engagement

- Selected studies



Enjoyment

+ve

- A test publisher found 94.3% of ppts (N = 1747) reported enjoyed playing a GBA
- Another test publisher found 90% of candidates feel that GBAs are the same or better than traditional assessments

-ve

- Candidates value ease of use and usability more than enjoyment. Most candidates would prefer job relevant test (e.g. work sample) over fun games. (Laumer et al. 2012)

Enjoyment mediated by individual differences:

- Oostrom et al (2011): candidate perceptions positively correlated with personality traits of Openness and Agreeableness

sten10
experts in people assessment

Face Validity / Engagement

- Selected studies

Gaming Expertise

A test publisher (2014) found 80% 'enjoyed' gamified learning tool BUT 'hard-core gamers' disengaged. Millennials most likely to logon, but quickest to drop out. Also found males more likely to engage with the game

Technology

Preuss (2017) found that 60% of candidates prefer using Gamified SJT over a traditional SJT.

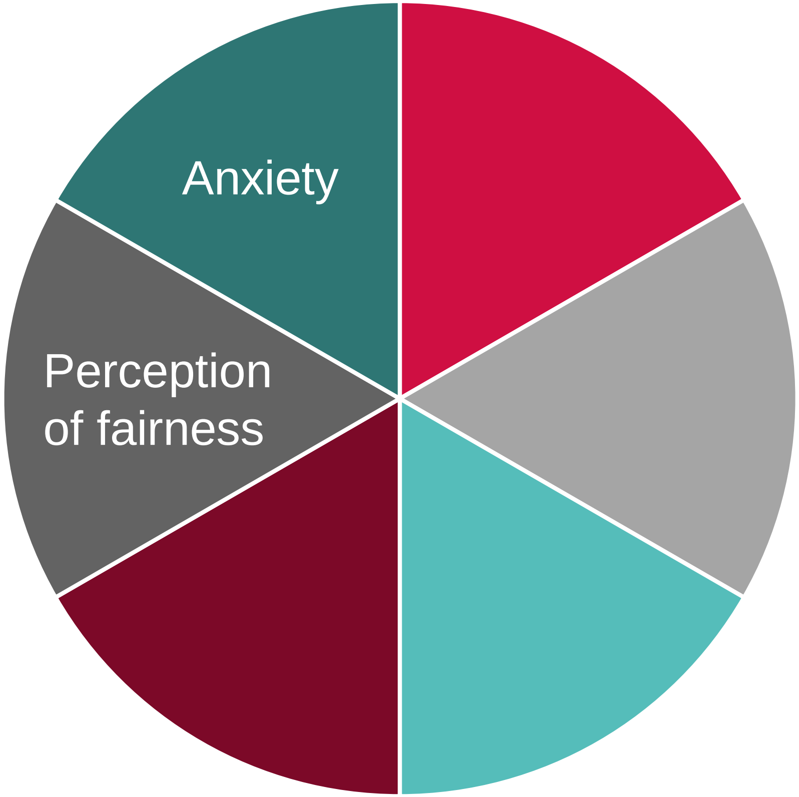
However, technological difficulties for some candidates resulted in lower perception of gamified SJT

sten10
experts in people assessment

Face Validity / Engagement

- Selected studies

Face Validity / Engagement
- Selected studies



Perception of 'fairness'

- A quarter of candidates believe completing an assessment on a mobile device would provide a 'fair' testing experience (Fursman & Tuzinski, 2015)
- Landers (2017) found test takers consider GBA 'fairer' than general cognitive ability tests
- Different publisher's manual showed 40% saw it as more fair, 40% less fair

Anxiety

- 74% (n=200) felt less anxiety for GBA, 89% enjoyed the selection process, 81% felt more excited about the prospect of working for the firm (test publisher research)
- Geimer et al (2015) found Candidates experienced higher levels of anxiety when feedback is given in game

sten10
experts in people assessment

Construct Validity

-Selected research

Big Five Personality

Van Lankveld (2011) 275 individual metrics in 'Neverwinter Nights' and found 1,375 correlations with Big 5 traits. However, some of these could be spurious. (n.b. n=44)

Short et al (2017) found no links to Big 5 using World of Warcraft. Fairly consistent support for preference for virtual teamwork and technology readiness.

Construct Validity

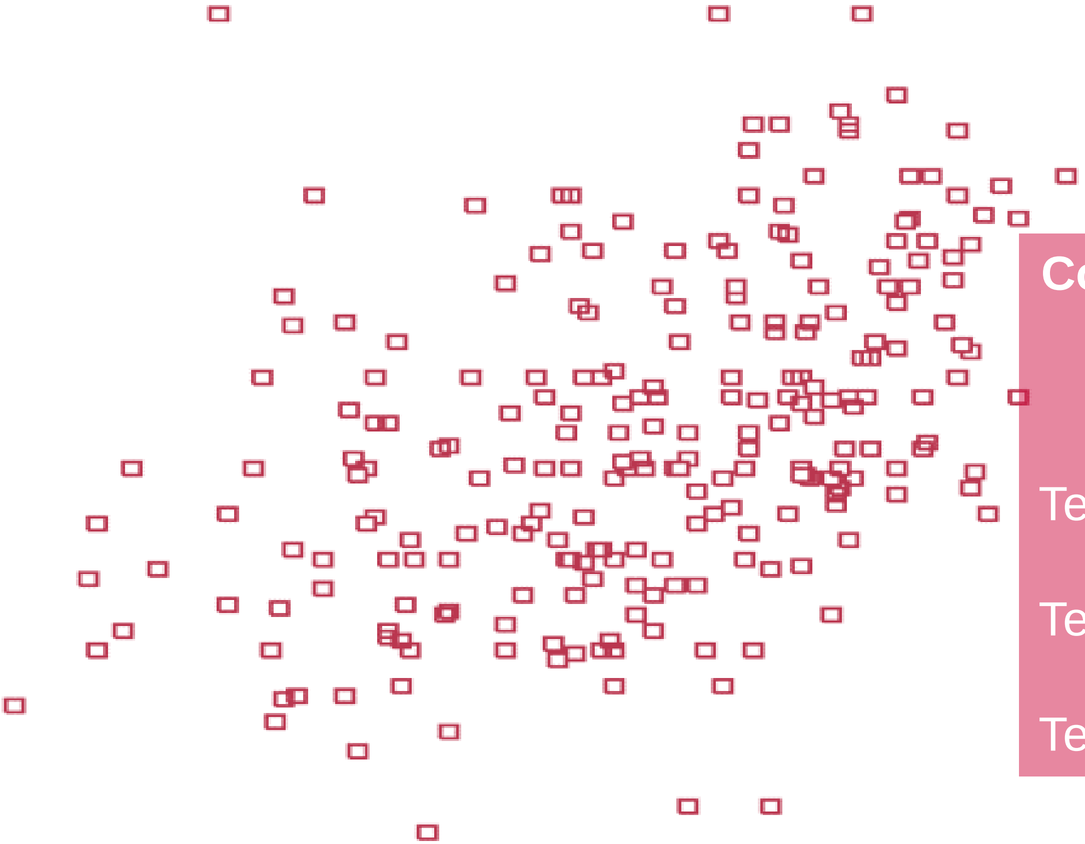
-Selected research

Working Memory/Fluid Intelligence

Baniqued et al (2013) found performance on games that required working memory and reasoning significantly correlated with performance on working memory and fluid intelligence tasks.

Construct Validity

-Selected research



Correlations with established measures of same constructs:

Test provider 1*: 0.24 to 0.44

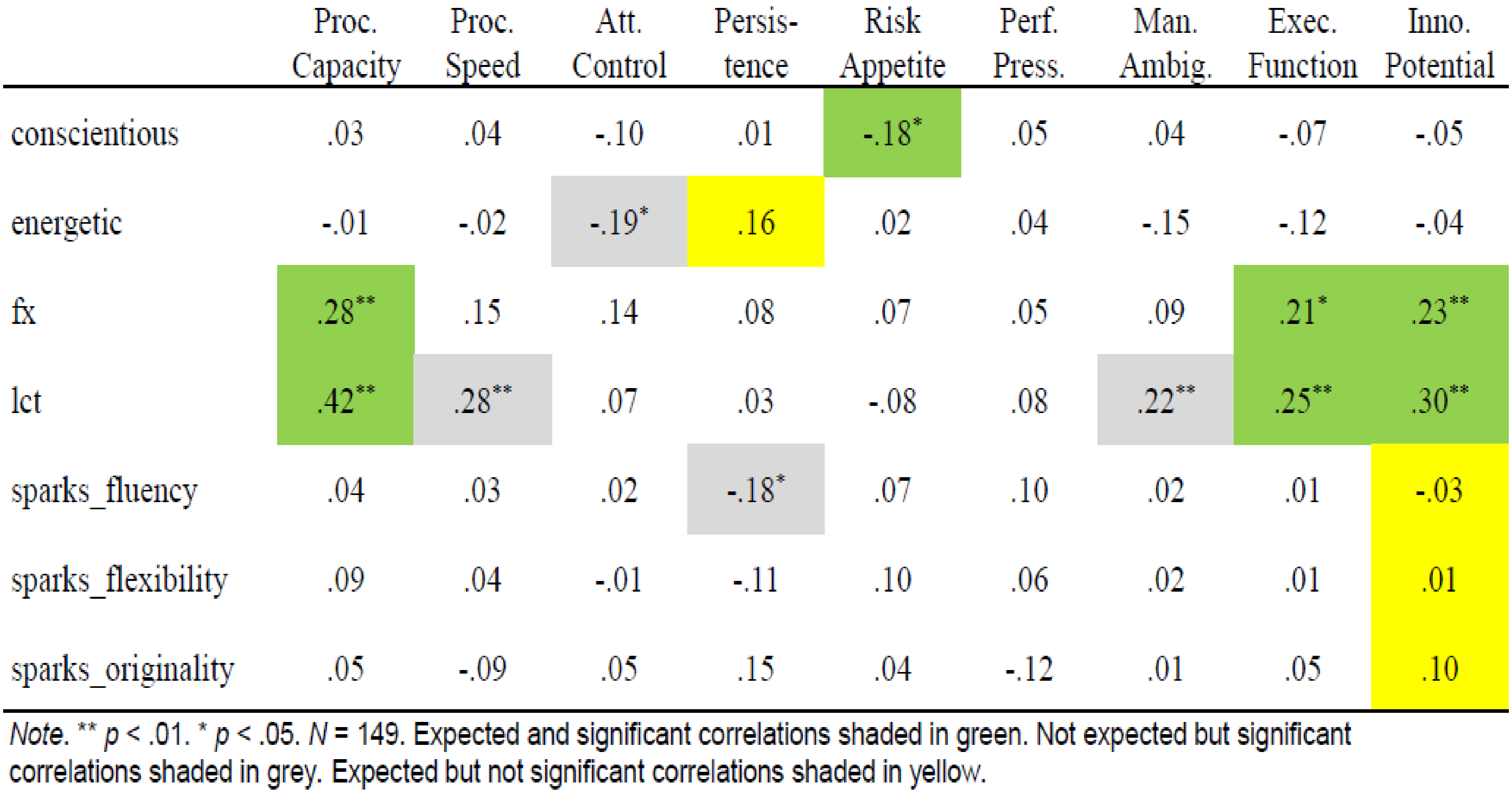
Test provider 2*: 0.2 to 0.26

Test provider 3*: 0.3 to 0.54

experts in people assessment

Construct Validity cont.

Figure 1 below for results. Personality constructs were found to be partly similar. There were varying results for cognitive abilities (divergent – different, convergent – similar).



	Proc. Capacity	Proc. Speed	Att. Control	Persis-tence	Risk Appetite	Perf. Press.	Man. Ambig.	Exec. Function	Inno. Potential
conscientious	.03	.04	-.10	.01	-.18*	.05	.04	-.07	-.05
energetic	-.01	-.02	-.19*	.16	.02	.04	-.15	-.12	-.04
fx	.28**	.15	.14	.08	.07	.05	.09	.21*	.23**
lct	.42**	.28**	.07	.03	-.08	.08	.22**	.25**	.30**
sparks_fluency	.04	.03	.02	-.18*	.07	.10	.02	.01	-.03
sparks_flexibility	.09	.04	-.01	-.11	.10	.06	.02	.01	.01
sparks_originality	.05	-.09	.05	.15	.04	-.12	.01	.05	.10

Note. ** *p* < .01. * *p* < .05. *N* = 149. Expected and significant correlations shaded in green. Not expected but significant correlations shaded in grey. Expected but not significant correlations shaded in yellow.

Criterion Validity

- Selected Research

Landers (2017) aimed to validate a cognitive ability GBA through comparison with a traditional test battery and found:

- The game predicted ‘grade point average’ outcome measure better than 15 separate Spearman’s g measures (Spearman’s g provided no ‘unique’ prediction).

Other case studies from GBA publishers:

- Prediction of selection success for air traffic controllers (2017). Significant difference between successful and unsuccessful applicants’ mean scores on GBA (p>.001)
- Overall AC pass rate in 2016 = 24% Now in 2017 = 40% (60% for some Business Areas)
- Hi / low manager rating versus GBA performance: 0.019 sig.
- Global Tech Co.: Quality of Hire survey: .162 and .220
- Prediction of competency scores in AC for sales roles ranged between .135 to .347.
- Prediction of competency performance at a retail company – Multiple R .539
- High performance contact centre agents made 66% more bookings in value than the lowest performers, 10% more calls in a month on average

Adverse Impact

Case study 1 (2016): 5,000+ participants, no adverse impact for:

Age, Gender, Ethnicity, Disability (after WM adjustment for dyslexia), Gaming experience, Handedness, Screen size

Case Study 2 (2017): 1,054 candidates, no adverse impact for:

Age, Gender, Race

Case Study 3 (2016): 155 participants, no gender differences on: “cognitive style”, “information processing competencies”

Case Study 4 (2018): No gender differences on personality responses

sten10
experts in people assessment

BUT, SHOULD there be group differences to reflect what we know about human nature?

Level 3
Conclusions

sten10

experts in people assessment

Summary

‘The practice of gamification has far outpaced researcher understanding of its processes and methods’ (Landers et al, 2015).

- Relative lack of peer-reviewed, academic (non-vendor-led) research.
- Of the evidence there is, reliability (internal consistency and over time), engagement and adverse impact data looks promising. Construct validity and parallel form reliability is positive, with caveats. Validity on later-assessment stages and on the job looks good, although more academic-led research would be beneficial.

Thank you!

Any Questions?