Gamification of psychological tests: three lessons learned

Philipp Sonnleitner

In an ideal world, test takers would love to engage with psychological and educational tests as much as their creators do. Being highly motivated, having fun and enjoying the task at hand would support people in bringing out their maximum cognitive performance which is – validity concepts of typical performance aside – what test developers really want to grasp in most of the cases.

A quick look in any psychologist's test closet, however, gives the impression that assessment should be as fun-free and laborious as possible. Some of the most iconic task types of psychological tests, for example Raven's famous matrices (see Fig. 1), appear intimidating at first and boring at best, exerting a certain fascination on puzzle enthusiasts or nerds, only. This state of affairs is not surprising though: the majority of cognitive tasks originated in experimental settings, which usually aim at minimizing emotional reactions of participants. Precise measurement instruments traditionally are not supposed to be fun (there is no entertaining clinical thermometer either) – quite the opposite is true: they should convey a certain seriousness of the assessment situation. In addition, since those cumbersome tasks served reasonably well in measuring and predicting people's abilities and characteristics - should we even bother with their appearance? According to recent debates on gamification research, we should.

In a nutshell, gamification means either disguising existing and validated assessment instruments as games (by introducing certain elements, such as appealing graphics and sounds, a narrative, and most importantly feedback) or using (computer-based) games themselves as valuable sources for psychological indicators. The (implicit) promise of introducing game mechanics to psychological assessment is that test takers actually have fun during the process, thereby forgetting about the fact that they are tested. Intrinsically motivated through play, they may be more likely to retrieve their highest potential or show their "true" characteristics when they are completing different tasks.

Ten years ago, being inspired by the then new hype on gamification, our lab was among the first to integrate game elements into a psychometrically validated test of complex problem-solving for the educational context: the Genetics Lab. Probably due to being at the height of the time by including game-like features and our open-access approach for test publishing, the Genetics Lab was featured in Psychology Today. It was downloaded more than a thousand times, was present on all continents except Antarctica, and in countries ranging from Austria to Zimbabwe. Originally published in English, German, and French, it was later translated into Italian, and Mandarin and was used in numerous studies.

Recent review articles on the topic, however, show that some of the bold claims related to gamification are only partially supported (Dichev & Dicheva, 2017, Lumsden, Edwards, Lawrence, Coyle, & Munafò, 2016). Game mechanics indeed seem to make the assessment process more enjoyable and they tend to increase test motivation. But their impact on task performance seems not so straightforward with few available, systematic studies finding mixed effects. Consequently, despite well-earned merits, some authors (e.g. Dichev & Dicheva, 2017) already see gamification on the descending branch of Gartner's technology hype cycle, and suggest developing expectations that are more realistic by sticking to a systematic research program on single game mechanics and their impact. From our own experience with the gamification of a cognitive test, we would like to share some insights that should be noticed and may help with this greatly needed, rigorous research scheme on gamification.

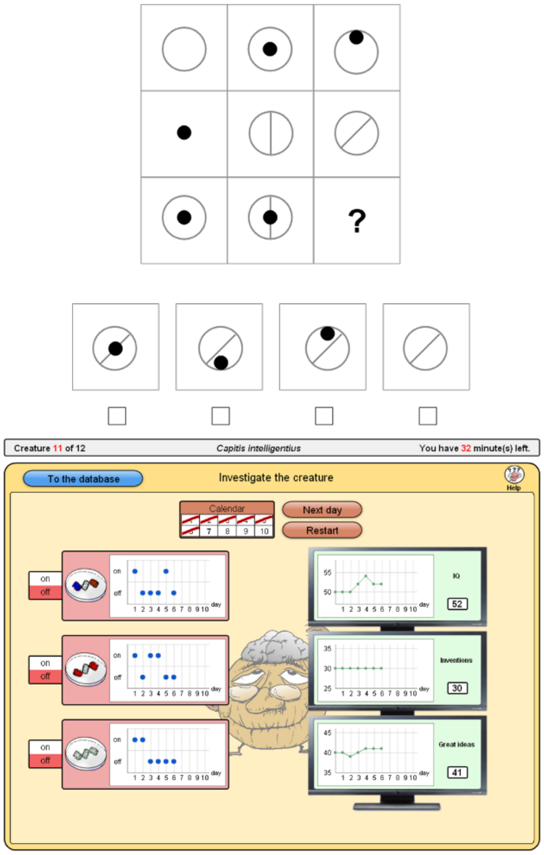

Figure 1: A typical matrices task (above) compared to a gamified computer-based item (below)

Lesson 1: Prepare for extended development cycles and join forces with experts.

In 2009, complex problem-solving (CPS) scenarios – although psychometrically valid – were, not surprisingly, complex in their very nature and still clumsy. They were complicated-looking computer programs that were used to assess university students' exploration and planning behavior. With the importance of CPS as 21st century skill on the horizon, it became clear however, that this assessment framework had to be adapted to the educational world by making it accessible for everyday teenagers of all ability ranges. Moreover, the often observed lack of students' test motivation in low-stakes educational assessment, especially within certain age and ability ranges, made us consider implementing game elements in order to elevate students' commitment. This shift, however, required drawing on concepts of multimedia learning to design the introduction and numerous, small-scale usability studies to make sure that students understood their tasks and that game-like mechanics didn't interfere with the assessment itself. In short, making a computer-based assessment look like a game meant greatly extending traditional test development cycles in terms of both time and resources. Finally, the Genetics Lab incorporated a storyline of a young researcher starting out in a scientific lab, which was supposed to be engaging for students. A comic-like design of the whole user interface (e.g., buttons and creatures) supported this narrative. After each task, students got feedback on their performance (1 to 5 stars). It paid off: 50% of the students indicated that they enjoyed working on the task and they even would love to repeat the 35 minute long test with new scenarios to explore.

Frankly, in 2019, with half of US households owning a dedicated game console (Nielsen, 2018), and almost everybody carrying a gaming device in the form of a smartphone in their pocket, these measures won't do it. Today's test takers immediately spot whether the coins they earn when solving a task are a simple motivational carrot or an integral part of a game. Thus, when studying effects of gamification, assessment researchers should team up with game designers and developers right from the start to get valid results. Additional resources should be secured for Usability testing or User experience (UX) design, making sure that test takers interact smoothly with the interface and extraneous cognitive load is low. Consequently, research on gamification, if taken seriously, will cost money and time, and requires the willingness to work in an interdisciplinary team.

Lesson 2: Gamers are different, boys are too.

When gamifying tests or using distinct games to assess abilities, you need to keep track of the gaming history of your sample. Before working on the Genetics Lab, we asked a representative sample of n = 563 students whether they were playing computer games and if they did, they were asked to specify which ones. Whereas gamers did not differ in reasoning ability from their peers, they showed lower grade point averages ($d = 0.25$, $p < 0.01$) boldly confirming common stereotypes. However, they shone on the

Genetics Lab's performance scores, with a much more detailed knowledge on the problems they had to explore (d = - 0.27, p < 0.01) and a substantially higher ability to solve these problems (d = - 0.45, p < 0.01).

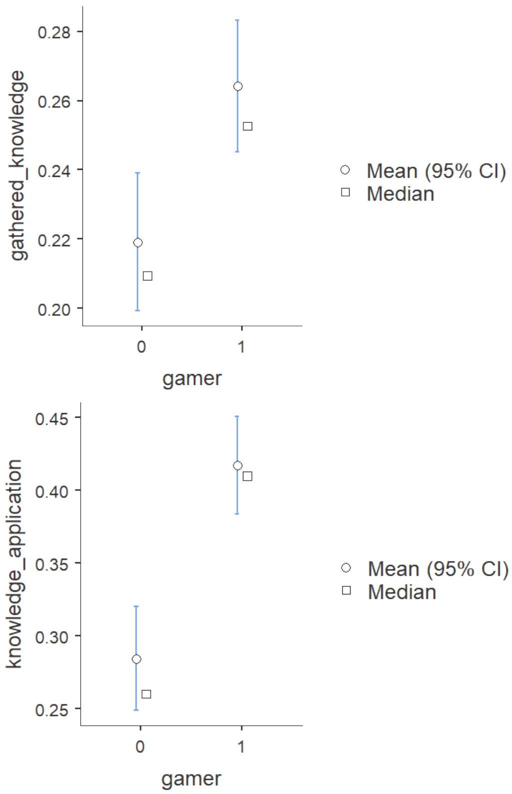

Figure 2: Difference between gamers (n = 338) and non-gamers (n = 225) on performance scores of the Genetics Lab

Crucially, this difference was not due to higher effort put into the test (d = -0.05, p < 0.54). Digging further into the data revealed that this significant difference was caused solely by gaming boys, who again, did not differ in reasoning ability, but apparently had much better complex problem-solving skills. Compared to gaming girls who preferred social simulation games (e.g. the Sims), boys listed action games (Jump & Run, Ego-shooter) or real-time strategy games as their favorite genres for killing time.

These results clearly do not warrant causal inferences, but they do give important hints on where to look or what to consider when further introducing game-mechanics to tests, or especially when using games to assess certain abilities. Do certain game elements trigger specific gamer populations because they are more common or known in certain gaming genres? Is the advantage of gamers in gamified cognitive tests due to the mode of testing (e.g. training effect) or really due to an underlying ability? Considering Multitrait-multimethod designs during psychometric evaluations of a gamified test would help solving this validity puzzle. Above all, measurement invariance with special regard to gender should be assured.

Lesson 3: Good ol' non-gamified Matrices aren't too bad

After having students work on the Genetics Lab, as well as on an old-fashioned paper-pencil Matrices reasoning test, we asked them which one they liked better. In fact, plain, abstract Matrices were preferred by all students except by gaming boys - specifically, those individuals who were outperforming the others on the Genetics Lab.

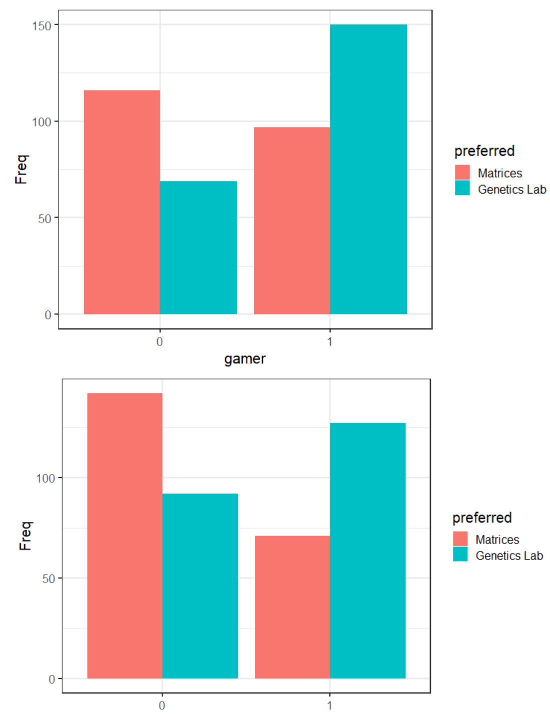

Figure 3: Preferred test of non-gamers (0, n = 225) and gamers (n = 338) and of girls (0, n = 234) and boys (n = 198)

However, equal effort was put in both forms of assessment with gamers showing generally a slightly higher value. We further asked students immediately after the instructions of each test about their expectation for test motivation and anxiety. Remarkably, students expected to have more fun and to experience less anxiety when working on the Matrices than when working on the Genetics Lab. This changed, however: when we asked students immediately after finishing the test, all differences vanished. Thus, it was the Matrices test instruction and not the gamified computer-based scenario that calmed fearful students and tricked them into thinking, psychological testing is fun.

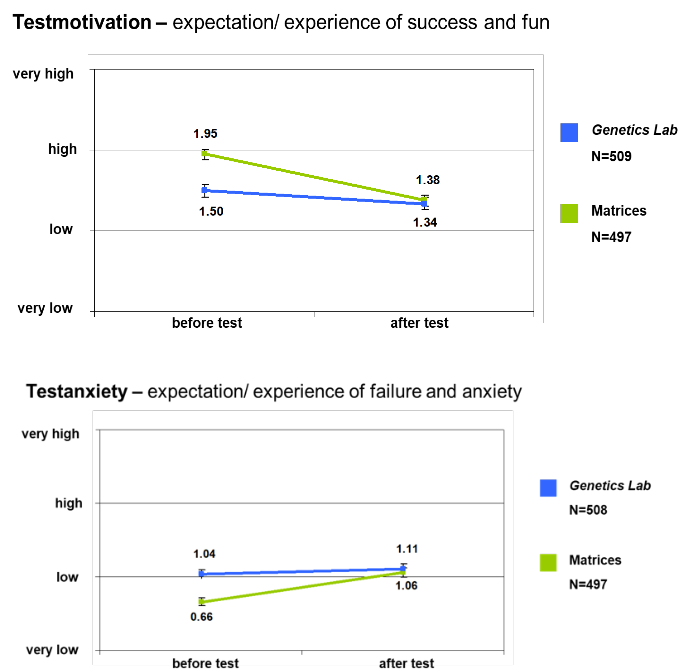

Figure 4: Comparison between Genetics Lab and Matrices concerning test motivation and test anxiety immediately before and after taking each test

Conclusion:

Gamifying psychological tests and using games for assessment undeniably has a lot of potential and attracts not only test developers but especially game-affine test takers. The hype on gamification probably also originates in the gaming biography and fond youth memories of today's test developers. The transition to more complex computer-based assessments additionally invites mimicking game mechanics. This fascination should not blind us to the fact, however, that developing well-gamified cognitive tests is a lot of work, inherits the danger of differentially affecting test takers, and that we should keep our expectations in check concerning its effects on increasing test motivation and allay fears. It is time for a well-funded, systematic research program to systematically explore gamification's potential and debunk some myths related to it.

References:

Armstrong, M. B., Ferrell, J. Z., Collmus, A. B., & Landers, R. N. (2016). Correcting misconceptions about gamification of assessment: More than SJTs and badges. *Industrial and Organizational Psychology*, 9(3), 671-677.

Dichev, C., & Dicheva, D. (2017). Gamifying education: what is known, what is believed and what remains uncertain: a critical review. *International journal of educational technology in higher education*, *14*(1), 9.

Nielsen (2018, February 1). The Truth is Out There: Knowing the Real Addressable Universe Is Crucial to TV Buys. Retrieved from https://www.nielsen.com

Dr. Philipp Sonnleitner
Research Scientist

Luxembourg Centre for Educational Testing
(www.lucet.uni.lu)
University of Luxembourg