Game-Based Assessment: TwoPractical Justifications

Thomas E. Heinzen
William Paterson University
Science Hall 2057
Wayne, NJ 07470
heinzent@wpunj.edu

ABSTRACT

Game-based assessment (GBA) is the application of principles of game design to measure performance when people are striving to perform at their best. Two simple but counterintuitive justifications suggest that GBA is a promising approach to assessment. First, people love to be assessed; they just don't like the ways that we have been doing it. Game designs cantake advantage of our impulse to make social comparisons to help motivate performance during assessment. Second, going to college is already a game. It has many of the features of a game such as points, a goal state, levels, awards, and leaderboards. Consequently, creating game-based assessments is an emerging skill set that needs a great deal of psychometric development as well as the knowledge and sensibilities of game designers.

Keywords

assessment, social comparisons, feedback loop, fear of failing, gamification, game-based.

1. INTRODUCTION

Game-based assessment (GBA) is the application of principles of game design to measure human performance when people are striving to perform at their best. If this is your first exposure to GBA, then keep three things in mind: (1) game-based *assessment*is focused only on assessment, not on finding new ways to teach and learn; learning is often a welcome, mastery-based byproduct of GBA. (2) Game-*based* assessment is not a video game; the internet makes GBA scalable but we are not trying to turn higher education into a giant game of *Jeopardy*.(3)*Game*-based assessment means that the assessment itself must be engaging and voluntary in order to capture peak performance and is probably the greatest (and most appealing) design challenge to GBA.

Interest in GBA appears to be approaching a critical mass that is starting to attract interest and funding [1, 2].It relies on the traditional principles of quality assessment [3] and blends the insights of game designers [4]. It is a mash-up that promises to make assessment more authentic.

2. TWO JUSTIFICATIONS FOR GBA

Although GBA can be used for summative assessment, its natural strength is monitoring performance by assessing the information trails that learners naturally leave behind when playing a game: formative assessment. In a digital GBA, that information trail can consist of four types of observations: TAPAS (or PASTA): Time to respond, Accuracy of answers,Points earned, and number of Attempts; best of all, all that information is sent automatically to a Spreadsheet for instant analysis and potential feedback. Two observations that have been hiding in plain sight help justify developing such GBAs: (1) people love to be assessed, and (2) the college experience is already a game.

2.1 Justification 1.Social Comparisons: People Love to Be Assessed

People love to be assessed – we do it all the time and the way we do it is through social comparisons. For example, we makean implied social comparisonevery time we glance in a mirror, compareresearch impact numbers, or feel envious that we are not working at a more prestigious institution. Making social comparisons is how we automatically gather important information about the self[6, 7] and GBAs take advantage of that impulse through three common game elements that comprise a feedback loop: points, leaderboards, and badges (PBLs). Game designers and educators tend to use these game elements in very different ways.

For example, game designers report points by creating an automatic or slightly delayed feedback loop that deposits a rich information trail during play.In basketball, you know how many points you have added the moment the ball goes through the hoop; in Tetris, every added layer of geometric blocks earns you more points – and you know it. It is a predictable, constructive feedback loop. By contrast, professors that return tests results promptly may be somewhat rare and they probably do not think very deeply about the consequences of how the timing of that information influences immediate learning or future performance. Adjusting our feedback loop demonstrates just one way in which game-*based* assessment does not have to be a game.

Another difference is thata game-designer uses points as rewarding motivators; educators tend to use points as punishment. For example, most games start with zero points and require player to earn their way to higher levels. By contrast, traditional testsoften take away points for every wrong answer. Game designers also use badges to offer a more formalized recognition of achievement. They may create adiversity of badges that have been thoughtfully designed and carefully game-tested. We educators tend to withhold themost motivating badge of all, the diploma, until the very last possible moment – just when its ability to influence student performance has ended. By contrast, game designers carefully sprinkle the pleasures of success throughout the game experience in order to lure players into deeper levels of their game. Educators have a lot to learn from game designers.

Leaderboards also can be used strategically and we need more basic research to know what kinds of social comparisons and leaderboards are most motivating or demotivating. Fortunately, game designers and social psychologists in particular already understand how and whyreward systems can backfire.For example, the over-justification effect tends to undermine the intrinsic motivation that is a necessary element for creative performance [8, 9]. In short, be cautious about rewarding students for what they already love to do;GBAs provide a path through that psychological minefield.

For example, a fully automated GBA can produce intrinsically motivating information, such as a (confidentiality-protected) two-dimensional scatterplot-leaderboard that changes each week by monitoring the relation between hours spent studying in the game environment and earned test points. The constantly changing scatterplot/leaderboard lets students and professorssee for themselves whether hours spent studying within the game environment really pay off. Such visual feedback is more informative and probably more motivating than hectoring students to study harder – especially when students are empowered to make their own data points move from one week to the next.

2.2 Justification 2: College is Already a Game

GBA recognizes another social truth also hiding in plain sight: *College is already a game.*Play only becomes a game when we introduce rules, goals, and arbitrary obstacles – such as 120 credits to graduate or specific course requirements. Onecritical difference between the experience of higher education and playing other games is how each deals with failure. In academic culture a failing student is threatened with a lower grade and dismissal if it continues; a failing department is threatened with a hiring freeze or dissolution; a failing university is threatened with loss of its accreditation. Those harsh consequences all inspire the kind of fear of failing that promotes risk-averse decision-making or a desperate fear-informed quality of achievement [10]. However, the experience of failure can be positively motivating; GBA uses many small failures as information goads that guide achievement.

Table 1. A qualitative comparison of responses to academic failures and game failures.

Academic failures lead to…	Game failures lead to…
summativeself-assessments	Formative skillassessment
a generalized fear of failing	Specific achievement desire
frustration/discouragement	Frustration/renewed effort
lower intrinsic motivation	higher intrinsic motivation
risk-averse decisions	exploratorydecisions

Table 1 compares the consequences of failing academically and failing in a traditional game. In the large college game, the drop out rate hovers around 50%, often the result of many small individual and system failures in many domains. However, the rate of small failures in games is even higher – about 80% [11].In poker, six out of seven poker players will lose the hand; in Tetris, every participant is guaranteed to fail. Almost all sports clubs fail to achieve their goal of winning a championship. Why is failing academically so intolerable while failing at games is frustrating but motivating and informative?

A surprising finding about student procrastination suggests a key difference between failure in academics and failure in a game: student procrastination is positively correlated with self-compassion; being kind to yourself[12, 13, 14, 15]. What do you think many students actually do when procrastinating? Klassen and Kuzuci [16] found that academically procrastinating adolescent boys in Turkey turn to computer games. These games provide feelings of being challenged in ways that they do not experience when studying [17]. In short, many procrastinating students are not running from a challenge; they are running to a challenge. They don't seem to be afraid of failing; they just experience academic failure as demotivating.

What would happen if we treated academics like the game it already is by reducing students' fear of failure and increasing the opportunity to play difficult,sophisticated games? In a study of 1,492 adolescents (50.8 % female), over the four high school years, Adachi and Willoughby discovered that more strategic videogame play predicted higher self-reported problem solving skills over time than less strategic videogame play [18]. Furthermore, strategic videogame play predicted higher self-reported problem solving skills that, in turn,predicted higher academic grades. In perhaps the most striking contrast to the fear of failing, game designers use the term 'fiero', Italian for pride, to describe how gamers feel when they overcome a difficult challenge [11]. Imagine an academic world in which fearless students are mildly addicted to the pleasures of being academically assessed!

3. GBA IN PRACTICE – AN EARLY REPORT

My first experience in game-based assessment was in an undergraduate statistics class whose schedule had been upended by what is now called Superstorm Sandy. Many students were without power for two or more weeks and the university was closed for three nights of this one night per week class. The storm and my interest in GBA allowed me to apply just few game mechanics to the class: Leveling up, fat points, and choice. Assessing their effectiveness was, of necessity, a qualitative endeavor that nonetheless suggested some insights worth testing more systematically.

3.1 Students Understand Leveling Up

Trying to speed-lecture through a 4-credit course with a one-credit lab didn't seem promising, so I stopped lecturing entirely. I turnedthe remaining course into a lab-based experience that focuses on executing and understanding the commonly used statistics program SPSS. Students could earn assessment points every class by conquering, for example, descriptive statistics, visual displays of data, *t*- tests, and so forth. But they could not go higher until they demonstrated competence in the lower levels, a process that gamers recognize as "leveling up" and educators think of as mastery learning.

Leveling up is arduously earning your way to the next level of accomplishment and game designers frequently use it as a motivational tool. For this situation, Level 1 was demonstrating capability in a series of exercises related to descriptive statistics, including some simple data transformations (reversing scores and compute statements). Level 2 was demonstrating competence at identifying and creating appropriate graphs – something that required and reinforced the material in earlier chapters. Level 3 involved mastering the three different *t*tests. The important point is

that students understood that leveling up presented a conquerable challenge with levels that would have to be repeated as necessary. In contrast to my initial worries, leveling up was an easy sell because it was a familiar dynamic expressed in the game language of what has become almost a common tongue among students. The informal contract between professor and student was well-understood.

3.2 StudentsSeem to Like "Fat" Points

One of the game mechanics that made leveling up work was what we can think of as "fat" points. You may recall that the earliest video games (e.g. PacMan) involved fat pointsas did their predecessors called pinball machines. The whirring, dinging, bouncing display of numbers somehow felt satisfying. Casino games employ the same strategy.

For this course, the original grading scheme was thrown off completely by the storm and students had to rely on my good will and assurances of fairness as we learned the material. It may have been the context of this class, but students responded eagerly to the repetition because (as they explained to me) the repetition "makes it easier and easier to earn points." They were mastering the material and having fun doing so, partly because thoseundefined, fundamentally meaningless point totals were rising so rapidly.

Fat points seem to be irrationally motivating because it should not matter whether a student earns 9 out of 10 possible points, 90 out of 100 possible points, or 900 out of 1000 possible points. But 900 seems to be experienced far more positively than 90 and it will be interesting to discover where a ceiling effect might kick in. More importantly, I found myself assessing students who were having more fun while striving to perform at their best in a class that some had entered with trepidation. They liked fat points but had no idea that I had introduced a game mechanic into the course.

3.3 Play is Voluntary

The irrationality of "fat" points appeared again when I asked two students after the course to devise a grading scheme for the course that they believed would be motivating and fair. They designed a point system that consisted exclusively of extra credit points – which amounts to the same thing as having no extra credit opportunities. But to them, it seemed to convey a greater sense of ownership; they were choosing how they could earn their points. In this case, the game mechanic of playfulness appeared to be influencing their perceptions.

Almost every definition of "play" emphasizes that it is voluntary behavior. Being a student is, technically, a voluntary behavior but the time lag between payment for and consumption of education produces what behavior economists call "payment decoupling." Payment decoupling explains why some students are irrationally happy when they learn that a class (that they have already paid for) has been cancelled [19, 20]. The expression of voluntary play that I introduced in this course was choice. The textbook we used had three sections of exercises: Clarifying the Concepts; Calculating the Statistics; and Applying the Concepts. Rather than assign exercises for students to complete, I used choice to convey a sense of voluntary participation. I told them that they could select any two exercises from each section. The unexpected positive outcome was that they scanned all of the exercises very carefully in order to determine which two they wanted to work on – and had to confront for themselves whether they wanted the easiest path through the course or the most instructive path through the course.

3.4 GBA Led to Unanticipated Positive Outcomes

Several unanticipated positive outcomes emerged from introducing these few game mechanics and using the associated GBAs. From my perspective as professor, embedding game-based principles created opportunities for more meaningful assessment. For example, the game mechanic of leveling up required substantial repetition. That repetition allowed me to feel more comfortable when I was obliged to assign some students a lower grade. I was convinced that they had had every opportunity to complete their work because the game design had allowed them to earn as many points as they cared to complete.

Another unanticipated positive outcome was that students helped one another learn SPSS and the concepts related to each statistical test. In the lab environment, students who I suspect never would have gone for tutoring requested and received help from fellow students who had mastered the material more quickly. There did not appear to be any stigma attached to struggling with the material because they were simply helping each other learn how to score more points in a game called SPSS.

One of the qualitative assessments of this course was essentially a focus group at the end of the class with an opportunity for students to anonymously comment on their experience. I probably should have been offended (see the underlined material) by some of their comments because I worked very hard and very late in each class in order to overcome the lost time due to Superstorm Sandy. But their comments made me wonder whether how potent the principles of game design might be.

- *Trial and error were good. <u>There was no real help along the way</u>, but there were no consequences.*
- *I like that we got to leave early if we learned all we needed to know for that week.*
- *I liked having someone other than the professor teach me; <u>it was much better</u>.* (!)

When I asked when, if ever, they had experienced the gamer's experience of an "epic win," students wrote:

- *When I began to be able to do the material and solve problems alone.*
- *Whenever I finished a new level I felt that I accomplished it.*
- *Completing the practice tests and doing them correctly.*

These comments suggest two outcomes with some searching implications: I was becomingunnecessary for learning and even at-risk were enjoying statistics.

4.SUMMARY AND CONCLUSIONS

Can GBA really help create a world of highly motivated students? The fit between game design and principles of backward design says yes [21]. We will have to redesign courses with a workableassessment plan at the start of the course design process sowe can know how we will know whether we are failing or succeeding. But such backward

designsmake the construction of GBAs much easier because each badge, achievement level, and grand quest can correspondto a specific learning outcome.

In most games and in most jobs, you get to make a few rookie mistakes as you get the hang of your new responsibilities; the critical assessment is not that you failed but whether you learned from your failure. Post-failure achievement is the more meaningful assessment we should be able to achieve through GBA.The promise of GBA is that we can measure performance when students are eagerly striving to be their best rather than frightened by failure.

Two features of GBA suggest that it is a practical approach to assessment. First, people love to be assessed because we use social comparisons to learn more about the self [7]. Second, going to college is already a game, and I suspect that many students have understood that far sooner than deans, professors, and curriculum designers. Our informal, qualitative assessment suggested that even a few, modest gamemechanics produced unanticipated benefits: Leveling up communicates a familiar expectation of challenge and useful repetition; fat points are irrationally motivating; and providing choice produced a sense of ownership and careful pre-evaluation of their choices.

Unlike traditional assessments, GBA makes failing informative and motivating. Birney, Burdick, and Teevan spent much of their careers studying fear of failure and the news, for the most part, isn't good [10]. But it is the fear of failing, not failing itself, that seems to produce counter-productive decision-making. People love to be assessed; they seek it out mostly through social comparisons. Higher education is already structured as a game. GBA applies the familiar, motivating techniques of game design to the urgent task of authentic assessment in higher education.

5. REFERENCES

[1] D. Ifenthaler, D. Eseryel, D., andX., Ge. X. *Assessment in game-based learning: Foundations, innovations, and perspectives.* New York: Springer, 2012.

[2] T. Kaya, A 'stealth assessment' turns to video games to measure thinking skills. *Chronicle of Higher Education, 57*(12), A13.,November 12, 2012.

[3] D. S. Dunn, M. A., McCarthy, S. C. Baker, and J. S. Halonen (Eds.) *Using quality benchmarks for assessing and developing undergraduate programs.* San Francisco, CA: Jossey-Bass, 2011.

[4] D. Dunning,C. Heath, and J. Suls,Flawed self-assessment.*Psychological Science In The Public Interest, 5*, 69-106.doi:10.1111/j.1529-1006.2004.00018.x, 2004.

[5] J. Schell. *The art of game design.*Boca Raton, FL: CRC Press, 2008.

[6] L. Festinger, A theory of social comparison processes. *Human Relations, 7,* 117–140, 1954.

[7] J. Suls, R. Martin, and C. Wheeler (2002). Social comparison: Why, with whom, and with what effect? *Current Directions in Psychological Science, 11,* 159-163, 2002.

[8] T. M. Amabile. The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology, 45*(2), Aug 1983, 357-376.doi:10.1037/0022-3514.45.2.357, 1983.

[9] E. L. Deci, R.,Koestner, and R. M. Ryan, "A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation". *Psychological Bulletin, 125:* 627–668, 1999.

[10] R. C. Birney, H. Burdick, H., and R. C.Teevan, *Fear of failure.* New York, NY: Van Nostrand-Reinhold Co. 1969.

[11] J. McGonigal. *Reality is Broken,* New York: Peguin Books. 2011.

[12] M. E. Deniz, Z. Tras, and D. Ayogan, D. An investigation of academic procrastination, locus of control, and emotional intelligence.*Educational Sciences: Theory and Practice, 9,* 623-632, 2009.

[13] M. Iskender, The influence of self-compassion on academic procrastination and dysfunctional attitudes. *Educational Research and Reviews, 6,* 230-234, 2011.

[14] K. G. Rice, C. M. E. Richardson, and D. Clark. Perfectionism, procrastination, and psychological distress.*Journal of Counseling Psychology, 59,* 288-302, 2012.

[15] L. J. Solomon, and E. D. Rothblum. Academic procrastination: Frequency and cognitive-behavioral correlates. *Journal of Consulting Psychology, 31,* 503-509, 1984.

[16] R. M. Klassen, R. M. and E. Kuzucu, Academic procrastination and motivation of adolescents in Turkey, *Educational Psychology, 29,* 69-81, 2009.

[17] T.Hainey, T. M. Connolly, M. H. Stansfield, and E. A. Boyle. The differences in motivations of online game players and offline game players: A combined analysis of three studies at the higher education level. *Computers and Education, 57, 2197-2211, 2011.*

[18] P. Adachi, and T. Willoughby, T. More than just fun and games: The longitudinal relationships between strategic video games, self-reported problem solving skills, and academic grades. *Journal of Youth & Adolescence, 42,* 2013, 1041-1052, doi:10.1007/s10964-013-9913-9.

[19] R. H. Thaler, .Mental accounting matters.*Journal of Behavioral Decision Making, 12,* 183-206, 1999.

[20] N. Wilkinson, N. *An Introduction to behavioral economics.*London: Palgrave MacMillan, 2008.

[21] G. Wiggins, and J. McTighe. Understanding by design, Upper Saddle River: Pearson Merill Prentice Hall, 2006.