Observations on the perspectives and limits of the evidence-based approach in the evaluation of gamification processes

Filippo Bruni

University of Molise, Italy, filippo.bruni@unimol.it, 0000-0002-5034-849X

Abstract

As continually greater attention is given to the processes of gamification, the dimension pertaining to evaluation must also be focussed on the purpose of avoiding ineffective forms of banalisation. In reference to the evidence-based approach proposed by Mayer and in highlighting its possibilities and limits, an experiment is herein presented related to teacher training, in which we attempt to unite some traits of the processes of gamification to a first evaluation screen. The data obtained, if they seem on the one hand, indicate an overall positive perception on the part of the attendees, on the other though, they indicate forms of resistance and of saturation with respect to both the excessively competitive mechanisms and the peer evaluation procedures.

Keywords: *educational game, evaluation methods, teacher training.*

Introduction: from game to gamification

Within the planning and management of training programmes, gaming as an instrument and a strategy goes quite back in time. The digital world has made possible a further expansion of the ludic universe and its educational use. Starting from the first phase of the videogame's evolution - generally understood as a game that implies a man–machine interaction in relation to an electronic apparatus equipped with a screen (Raessens & Goldstein, 2005, p. xii, Andreoletti, 2010, p. 83) - new ways of viewing the relationship in training terms between games and digital resources have been affirming themselves. In this logic, the process known as gamification is receiving ever more attention, that is, the use of a ludic approach in a context that is not at all ludic: 'Gamification brings together all the disparate threads that have been advanced in games for nongaming contexts' (Zichermann & Cunningham, 2011, p. xiv). The perspective that appears before us, beyond that of the fascination and the involvement that games have always held, is to understand the actual effectiveness of gamification processes in cognitive terms. The underlying idea is to take the inherent features of games such as high levels of involvement and apply them to contexts whose ultimate end is not playing, but learning. The difference between gaming and gameful activities (Staccioli 2008, p.19) can be viewed under this light: what characterises a gameful activity is the fact that the aim of the game, not being intrinsic to it, takes on an instrumental function.

The processes of gamification so cover an extremely vast area in which it is possible to include a multiplicity of other ludic possibilities. If, for example, we wanted to compare serious games - that can represent a type of game, (though these are usually called by other names such as edutainment, playful learning, edugaming, all of these wishing to tie gaming with the teaching/learning environment) - although they have common ends with the processes of gamification, they are to be distinguished from the latter by what can be described as a well-defined game space: 'Serious games tend to take the approach of using a game within a well-defined game space like a game board or within a computer browser, while gamification tends to take the use of a game outside of a defined space' (Kapp, 2012, p. 16). Moreover, to continue along this line of reasoning, it follows that 'Serious games are a specific sub-set of the meta-concept of gamification' (Kapp, 2012, p. 18).

As there is no more clean-cut separation between gaming seen as an autonomous activity with an end in itself and the working and learning dimension; the relationship between formal, informal and non-formal learning is undergoing a new conceptualisation: ludic mechanisms designed for non-formal learning contexts come to be intentionally used within formal learning paths.

Observations on the perspectives and limits of the evidence-based approach in the evaluation of gamification processes
Bruni

1. Defining the processes of gamification

The risk, perhaps because of the presence of such an extremely broad category, is that of simplification and banalisation given that the use of elements and mechanisms typical of gamification, when taken one by one, do not support an adequate approach. It has been justly observed that gamification cannot be reduced to badges, points and rewards: the automatic introduction of certain ludic elements represents a minimum level that can turn out to be ineffective. The processes of gamification achieve their maximum potential when a narrative dimension is introduced (Kapp, Blair & Mesch, 2014, pp. 103-115): the idea of flux in this sense is central (Zichermann & Cunningham, 2011, pp. 16-18), as it allows for the avoidance of the opposite extremes of boredom, on the one hand, and excessive anxiety on the other, thanks to a progressive involvement of the abilities of those participating in ever higher levels of play. All of this implies that the identity of the participants be put into play, thus generating significant learning (de Mul, 2005, Bruni, 2007, Gee, 2007, pp. 45-69). And so conceiving the processes of gamification, not simply for the ludic mechanisms used, but for the purposes to be pursued is logical: to involve and motivate. The definition of gamification proposed by Kapp appears to be quite exhaustive: 'Gamification is using game-based mechanics, aesthetics and game thinking to engage people, motivate action, promote learning, and solve problems' (Kapp, 2012, p. 10).

2. Theoretical framework: between visions and practical directions

The effort to understand and define gamification processes falls into the line of research that investigates the relationship between digital games and learning or teaching. Such contexts are not easy to reconstruct nor are their many complex aspects easy to identify.

It may be useful to follow Mayer's (2014, p. xiv) systematic examination of the research literature, which looks at the subject from three different perspectives. The first perspective deals with the attempt to offer a big picture through a range of edited volumes. Foremost among such works is Raessens and Goldstein's Handbook (2005), written with the primary intent of 'including the broadest range possible of perspective on gaming' (Raessens & Goldstein, 2005, p. xi), thus hardly covering such issues as teaching and learning. The second perspective is the one offered by the so-called 'visionary volumes', a range of works offering general perspectives on the matter: from Prensky's (2006) optimistic vision of the educational potential of videogames, to Alternated Reality Games (ARG) - an emerging typology of games whose main characteristic is the intertwining of physical and virtual reality (McGonigal, 2011) - to the singling out of learning principles related to videoludic activities (Gee, 2007). The third approach relates to 'practical guides' that specifically refer also to gamification processes (Kapp, Blair & Mesch, 2014).

The space between the indication of principles, however, supported by empirical examples and practical directions, has not been adequately investigated. Beyond the huge attraction and fascination that gaming has always been able to generate, the question arises as to how to gain a thorough understanding of the actual effectiveness of game and gamification processes as cognitive tools, also identifying the specific cognitive functions involved (Trinchero, 2014, pp.10-13).

3. Evaluating the processes of gamification: Mayer's proposal

As regards evaluation, Mayer's recent and very thorough perspective underlines how the evaluation dimension in relation to gaming and videogames has been neglected for far too long.

There are three dimensions, or better yet, pairs of elements to be compared and to which attention must be paid: value added, cognitive consequences and media comparison. The first (value added) is associated with the comparison between the basic version of a game and a more advanced version with additional and specific characteristics. The second (cognitive consequences) concerns the differences in cognitive skills between those who play a certain game and those who do not. The third (media comparison) evaluates the difference between using traditional methods and games in relation to the teaching of a specific content (Mayer, 2014, pp. 8-11).

From the methodological point of view, the importance of the use of control groups must be underlined, as must the random assignment of the components of the control groups, and the use of appropriate statistical measures too (Mayer, 2014, pp. 34-37), especially with respect to effectiveness (What works?). When the research moves towards other dimensions relative to context and the various types of users (When does it work?), to the means of functionality (How does it work?) and the description of the processes (What happens?), then methodologies that go beyond the strictly quantitative approach (interview, observation) (Table 1) come into play.

A critical point of Mayer's approach is given by the context. On the one hand, one can understand and easily share the intention a) to identify on a rigorously experimental basis the characteristics at the origin of effective training avenues, b) to replicate such characteristics in other cases. On the other hand, if the context in its complexity is taken into consideration, a strictly experimental approach could turn out to be insufficient a) in analytical terms b) in a construct that views to reuse training in the planning of new training programmes. Mayer himself, when taking into

consideration the context, a) indicates a limited and general series of criteria to be examined (subgroups, learning context, types of content) and b) proposes a 'quasi-experimental' approach recognising the difficulties in proceeding with rigorously experimental modes (Mayer, 2014, p. 30). The horizon that appears, beyond that of the approach proposed by Mayer, thus, is that of understanding and analysing in the most complete manner the articulated context in which the processes of gamification come into play.

Table 1. Four goals of game research. Source: Mayer 2014, p. 29.

What works?	Value added: Does adding feature X to a game improve learning? Cognitive consequences: Does playing game Y improve cognitive skills? Media comparison: Do people learn better from conventional media?	Experimental comparison
When does it work?	Value added: Are the effects of feature X stronger for certain kind of learners? Cognitive consequences: Does playing game Y improve cognitive skills better for certain kinds of learners? Media comparison: Do certain kinds of people learn better from games than from conventional media?	Factorial comparison
How does it work?	Value added: Why does adding feature X to a game improve learning? Cognitive consequences: Why does playing game Y improve cognitive skills? Media comparison: Why do people learn better from games than from conventional media?	Experimental comparison, interview, observation, survey
What happens?	What do people do when they play a game?	Observation, survey, interview

4. Teacher training and gamification: an experiment

In the light of the theoretical framework outlined above and the critical point there indicated, some gamification elements were introduced into a training programme; and by means of an online questionnaire/interview, the attendees were asked to evaluate the programme in general as well as specifically with regard to the gamification process introduced.

For purposes of the experiment, we used the course intended for the training of teachers of middle/secondary school under the auspices of the University of Molise. The attendees, all of Italian nationality, were prevalently female (55 of 86, i.e. 64 %) and covered a wide range of disciplinary areas for high school teaching: physical education and sport; agrarian chemistry and agrarian sciences; zootechnical studies and animal production sciences; food sciences; technical education; natural sciences; chemistry and microbiology; Italian, history and civics; history and philosophy; mathematics and physics; literary studies and Latin; business and economics, law and economics. Given the factors involved - the low number of hours (18), the high number of attendees (86), the availability of only one room with video-projection, the impossibility of activating several labs in a coordinated way - the simplest solution in a traditional vision of things would have consisted in classroom lessons with a final oral exam. Given, however, the possibility of using e-learning for 50 % of the time, we opted for a different set-up. By using a blog, we tried to propose the course as a game divided into four challenges, and so we tried to transform traditional assignments (for example, identifying didactic materials) into contests (who is able to find the best material?). The first is centred on presentation of one's own idea of the web, the second on the identification of digital resources to be used as teaching activities, the third on apps and videos for teaching, the fourth for course evaluation. The logic used was that of introducing a minimum nucleus of gamification elements: a) to increase the levels of involvement and motivation, b) to promote forms of interaction and peer evaluation among the participants, c) to render the course evaluation parameters transparent. Thus, four meetings were alternated as lessons and evaluation sessions, presenting the course contents as feedback to each of the above challenges. The module documentation is available online within the blog at http://proviamociesperiamobene.blogspot.it/.

As an example, here is the post published for the second challenge:

"Group 1 […] Second challenge: what resources for teaching?

As agreed during the lesson, the second challenge consists of:

1. Identifying a resource online that you consider may be used in your teaching activities. Any online resource is fine: institutional site, blog, Facebook group, software, repository of images and didactic material, videos, just as long as these are not commercial products. In a comment to this post, publish the link (Time: 30 minutes; Points: 1 out of 30). (N. B. This activity is to be considered obligatory.)

2. Briefly defend your choice: Why this particular resource? How do you plan to use it? (Time: 30 minutes; Points: 1 out of 30). (N.B. This activity is also to be considered obligatory.)

3. View and read all comments of the colleagues from your group […] and choose the most significant comment, briefly defend your choice by means of logical argumentation (Time: 30 minutes; Points: 1 out of 30). The deadline for publication of comments is set for midnight February 9th.

4. Only for the public competition coordinators […]: read the comments of your colleagues belonging to the same class of competition and communicate in a suitable post […] the comment which has received the greatest number of clicks (Time: 30 minutes; Points: 1 out of 30).

Information concerning extra points are as follows: a) 1 point out of 30 for the authors of those comments which each class of competition indicates as the most significant b) at the oral exam, the examiner may delve more deeply into the comments. The attribution of an additional 1 point out of 30 is at the discretion of the examiner. Good luck!"

The elements of gamification used are quite few in number: an adequate lexis to present the activities of the course as a challenge, the distinction between obligatory and optional levels, the definition of points that make the evaluation transparent, indications concerning the deadlines understood as rules of the game, an alternating between feedback and challenges such that there is a rhythm to the activities and no down time. In particular, the points mechanism has been used to promote interaction and peer evaluation.

5. Evaluation tools

As an evaluation tool, in relation to a) the course in its entirety, b) the elements of gamification introduced, c) some elements of the context, a questionnaire/interview with open questions was used as the methodology, in the awareness that this may constitute only a first step towards more structured forms of evaluation. On the one hand, given that the questionnaire/interview allowed us to determine the level of satisfaction of the attendees, Mayer's proposed scheme does not apply given that it separates quite cleanly the evaluation of learning and that of satisfaction (Mayer, 2014, pp. 57-58). On the other hand, it is possible to identify a series of elements that, though they belong to the area of the attendees' perception, touch upon some ambits identified by Mayer (cf. Table 1, How does it work? What happens?); these can be legitimately explored by means of interviews in a 'quasi-experimental' perspective. The brief questionnaire comprising only three questions was administered to each of the four groups attending the course with a specific post on the blog.

"Now to the last challenge. Perhaps the most difficult. Evaluating is not at all easy, though it is indispensable for taking decisions. If I find myself in the coming years in the condition of holding another similar course, I will have to decide what to keep and what to change. Having your feedback for me is important. I know full well that evaluations should be anonymous so as to allow for the utmost freedom in judgment. I believe that anonymous mechanisms will be activated. However, as I was telling you at class, in the Italian context, unfortunately, there are two types of behavior that oscillate between servile adulation in public and insults in the ever less private sphere. You are, however, adults and I do not think that I have any particular capacities in inhibiting your evaluations and I assure you that I appreciate the frankness of those who are able to express critical thinking. After all, the issue here is that of evaluating the course and not the professor (teaching means evaluating competencies and not people) and I hope that this will be for you an opportunity to reflect on what you've learned.
Three questions:
1. I would ask you to identify an aspect of the course, which you think should not be repeated. What is the critical point not to be repeated?
2. Which aspect should instead be repeated? Limit yourself to a single aspect.
3. As you have no doubt understood, I introduced certain elements of gamification into the course (not many and in a casual manner): a) Lexis ('challenge' rather than 'activity' or 'submitting'); b) grading (in the form of positive comments); c) rules (points, deadlines: in reality, these are aspects common to evaluation techniques). Help me to understand: Do you think that these were useful? Were they excessive? Or, on the contrary, could we have accentuated the challenge aspect?"

6. Results and discussion

With respect to the third question, in reporting the evaluations of those who took the course, three levels became apparent. The first is that of the positive reactions, that is, those evaluations that considered the introduction of gaming into the course a useful development. Those evaluations that signalled minimum critical points entered into this category and more precisely: the point assigned to the coordinator of each group, though it is seen as an elective task, upsets the evaluation nonetheless; the request for fewer challenges that could not be entertained, as the gamification aspect would have come to be missing; having more time available, for the purpose of reducing anxiety, but this too, at least within tolerable limits, is part of the game; problems with respect to the composition of the groups that, given the

ties to the class of competition, could not be resolved, if not with great difficulty. The second level concerns the partially positive evaluations, that is to say those that viewed the introduction only partially useful and the third those that considered it harmful. Table 2 shows how gamification overall was received positively.

Table 2[1]. The attendees' evaluations.

	Attendees	Gamification useful to learning	Gamification partially useful to learning	Gamification not useful or harmful to learning
Group 1	24	16	7	1
Group 2	22	20	1	1
Group 3	22	10	9	3
Group 4	18	11	6	1
Total	86	57 (66%)	23 (27%)	6 (7%)

We will here concentrate on the critical elements, wilfully ignoring the elements supportive of the usefulness of gamification.

An unexpected result was given by a nucleus of those attending the course who declared that they had minimal digital skills, so much so that the blog - a tool considered by many to be in its maturity from a technological point of view - was viewed as a novelty: the level of digital skills (inevitably linked to the practice of gamification) impacted on the ability to make good use of the technology.

The second datum links the lack of usefulness or harmfulness of gamification (introduced in any case in a limited way) not so much to specific modes of use, but rather to rejection with regard to competition and in some cases with respect to the very idea of competition per se ('We live in a world where competition is exalted and where coming second is not of any value' F.C: group 2), in others, by forms of saturation with respect to selection processes such as those necessary for accessing the training, which is based on points.

The third datum poses an even more general question and concerns the resistance towards evaluation mechanisms in general as well as to peer evaluation in particular. ('When I think of my training as a teacher, I always think of what I can do to become a better teacher […] and I wouldn't do any better depending on the number of points available' S.S. group 3). In this way, one of the strong points of gamification is brought into question, that is to say, the link with rigorous and transparent mechanisms, which can take on the guise of certification of skills acquired (badges, for example). The resistance to peer evaluation, though such a mechanism is quite commonly used in informal ludic environments such as social networks, is a function of the fact that one is the object as much as the subject of formal evaluation processes.

Conclusion

In conclusion, generalising the outcomes of the experiment, it can be observed that:

1. Factors such as the level of digital skills, the greater or lesser openness to competitive modes, diverging visions on evaluation in general and among peers in particular, all these factors contribute to the perception of utility of the processes of gamification, though probably on the levels of motivation and learning as well.

2. From a methodological point of view, there is the problem of how and how much a number of general variables connected to the context such as those just indicated, impact on specific outcomes obtained via rigorously quantitative methods within an evidence-based approach. Respect for the complexity and the diversity of different contexts means that an excessively rigid evidence-based approach could have difficulty in identifying all the elements that go into making a programme effective. Moderate forms of evidence-based approaches (Calvani, 2012, pp. 20–21) that are able to make use of qualitative and 'quasi-experimental' forms could positively integrate with ethnographic methodologies that, in relation to video games, have produced significant research (Ito, 2010, pp. 195–242).

3. From a planning or management view, the critical points identified indicate that the introduction of the processes of gamification or even of only some mechanisms attaining to gamification, cannot be accomplished a priori without adequately and carefully calibrating with respect to the specificities of the context and the attendees, failing which, the programme could be ineffective.

---

[1] The table above can be better understood by considering that the groupings were realised only to facilitate interaction and consequently, the disciplinary areas to which each attendee belonged were aggregated in order to form groups of more or less the same size. The relationship between the single disciplinary areas and the process of gamification activated thus was not examined and so it remains an area of study of particular interest, the disciplinary area of physical education being a good example.

References

Andreoletti, M. (2010). Il videogioco. Questioni, tassonomie, similitudini. *Rem. Ricerche su Educazione e Media*, II(1), 81-103.

Bruni, F. (2007). Identità in gioco, *Journal of e-Learning and Knowledge Society*, III(3), 29-37.

Calvani, A. (2014). *Per un'istruzione evidence-based. Analisi teorico-metodologica internazionale sulle didattiche efficaci e inclusive*. Trento: Erickson.

de Mul, J. (2005). The Game of Life: Narrative and Ludic Identity Formation in Computer Games. In J. Raessens, & J. Goldstein (Eds.) (2005). *Handbook of Computer Game Studies* (pp. 251-266). Cambridge (Massachusetts) – London: MIT Press.

Gee, J.P. (2007). *What Video Games Have to Teach Us about Learning and Literacy*. New York: Palgrave Macmillan.

Ito, M. (Ed.) (2010). *Hanging Out, Messing Around, and Geeking Out*. Cambridge (Massachusetts) - London: MIT Press.

Raessens, J., & Goldstein, J. (Eds.) (2005). *Handbook of Computer Game Studies*. Cambridge (Massachusetts) – London: MIT Press.

Kapp, K.M. (2012). *The Gamification of Learning and Instruction. Game-based Methods and Strategies for Training and Education*. San Francisco: Pfeiffer.

Kapp, K.M., Blair, L. & Mesch, R. (2014). *The Gamification of Learning and Instruction. Fieldbook.* San Francisco: Wiley.

Mayer, R.E. (2014), *Computer Games for Learning. An Evidence-Based Approach*. Cambridge (Massachusetts) – London: MIT Press.

McGonigal, J. (2011). *Reality is Broken: Why Games Make Us Better and How They Can Change the World*. New York: Penguin Press.

Staccioli, G. (2008). *Il gioco e il giocare. Elementi di didattica ludica*, Roma: Carocci.

Trinchero, R. (2014). Il gioco computerizzato per il potenziamento cognitivo e la promozione del successo scolastico. Un approccio evidence based. *Form@re*, XIV(3), 7-24.

Zichermann, G., & Cunningham, C. (2011), *Gamification by Design. Implementing Game Mechanisms in Web and Mobile Apps*. Sebastopol: O'Reilly.