Using Comparative Behavior Analysis to Improve the Impact of Serious Games on Students' Learning Experience

Dominique Jaccard, Jarle Hulaas, Ariane Dumont

School of Business and Engineering, University of Applied Sciences of Western Switzerland
Avenue des Sports 20, 1401 Yverdon-les-Bains, Switzerland

`firstname.lastname@heig-vd.ch`

Abstract. In the last decade, the use of serious games as a teaching and learning tool has steadily increased in many disciplines. Nevertheless, serious games are still facing crucial challenges, such as their integration in the global learning process. On the other hand, with the increased adoption of online applications and courses, it is becoming possible to collect and centralize large amounts of trace data generated by players. Such data may be used to produce statistics on students' behaviors inside pedagogical serious games, both as individuals and aggregated as groups (e.g., classrooms). In this paper we propose a classification of potential uses of statistics in serious games and give new insights into how statistical analysis of groups' behavior may impact positively on the learning process. We also present experimental results obtained during a large-scale game deployment using the Wegas platform, our open source platform for game authoring and execution.

Keywords: Serious games, Statistics, Comparative behavior analysis, Classroom orchestration, Teacher empowerment, Learning analytics, Competency management, Deep learning.

1 Introduction

The professional world is becoming more and more volatile, uncertain, complex and ambiguous. According to the World Economic Forum [25], students therefore need more than traditional academic learning: they must also develop new skills through education technology such as serious games (SGs). The use of SGs for pedagogical purposes has indeed become more widespread, from the health sector to engineering or management education, with some positive impacts reported (see e.g., [4]). Interactive teaching strategies have also proven to increase student attendance and engagement [7] and to foster higher performance [11]. But there are still many challenges to take up for effective acceptance of SGs in the educational process [12], such as how to promote deep learning [9] and trigger the higher levels of Bloom's taxonomy of learning [3] while integrating SGs in different teaching strategies.

R. Bottino et al. (Eds.): GALA 2016, LNCS 10056, pp. 199–210, Springer, 2016.
The final publication is available at https://link.springer.com/chapter/10.1007/978-3-319-50182-6_18

Since SGs are used more and more in networked environments, it is becoming possible to centralize trace data that reflect students' behavior while they are playing. This in turn enables the aggregation of datasets with different scopes. Analysis of such statistics may lead to the definition of strategies for improving the learning process. The activity of collecting, analyzing and reporting data about learners and their contexts is called *learning analytics* [10]. In this paper, we are mainly interested in tracking and comparing players' behaviors, in particular their decision making, as opposed to more synthetic scores such as examination grades.

Whereas learning analytics has been previously studied as a means to follow, assess and predict student performance [1][2][18], in this paper we argue that behavioral statistics will help address other important challenges that are preventing the larger adoption of SGs. More precisely, we propose to compare behaviors inside games by categories of players, such as by different years, backgrounds (e.g., management or engineering), levels (e.g., undergraduate or postgraduate studies), environments (e.g., academic or professional) and teachers. The comparative analysis of such statistics has the potential of impacting the learning process in a positive way while supporting different pedagogical approaches. The same principle may also be leveraged to obtain better real-life decisions in companies or organizations.

In the following section, we propose a classification of the potential uses of statistical behavior analysis, with a special focus on SGs with pedagogical objectives. Following this, we present a study that took place in the frame of a project management game, which we use to validate the idea of comparing behavioral statistics in SGs. Then we discuss the significance of these results as well as their limitations. The paper concludes with some perspectives on possible extensions and future research directions.

2 A Classification of Statistics in Serious Games

Many SGs already make their statistics available to players and teachers. A promising but lesser studied field of learning analytics is to determine which aggregation levels are relevant and to which stakeholders this information will make sense. The purpose of this section is to provide an overview of these questions.

2.1 Stakeholders of Serious Games in Education

Many stakeholders are involved in the use of SG for educational purposes, including not only students and teachers, but also educational program chairs, SG designers, pedagogues, educational quality managers and instructional designers. All of these stakeholders have specific interests and questions about the use and results achieved with SGs. Decision makers outside of the educational setting, like human resource managers, may also be interested in using SGs as a support for their decisions.

2.2 Levels of Data Aggregation

Trace data may be aggregated with different scopes to produce meaningful statistics. The following scopes seem to be the most relevant to students' learning experience:

1. Single game play: e.g., individual statistics about each student, such as the percentage of correct answers or decisions he or she made throughout the game.
2. Classroom session: e.g., global statistics on the distribution of choices made inside a game played in parallel by several participants, such as in a classroom.
3. History of sessions for a specific course: e.g., statistics on the same course throughout a number of years.
4. Sessions of different categories of players: e.g., statistics comparing classrooms with different profiles that have played the same SG.

2.3 The Stakeholder / Scopes View

The use of learning analytics in SGs has previously been studied along 3 axes: what, when and where to evaluate [18]. In this paper we focus more on the potential applications of data aggregations and therefore are more interested in the question of what aggregations are useful, as well as for whom and why data are collected and analyzed.

The following table suggests which data aggregations may be relevant to the different classes of stakeholders. It embodies the intuitive idea that increased scopes of data aggregation will be of interest to a larger number of stakeholders because the additional information will contribute to answering more questions. Additionally, in order to introduce some structure in the list, we distributed stakeholders into three levels (micro, meso and macro) according to the size of their primary population of interest, but the boundaries between these levels are not strict.

Table 1. The stakeholder / scopes view

	Micro level		Meso level		Macro level	
	Learner	**Teacher**	**Educ. program manager**	**Game designer**	**Re-searcher**	**Other decision maker**
1. Single game play	x	X				
2. Classroom session	x	x	x	x	x	x
3. History of sessions (same course)	x	x	x	x	x	x
4. Different categories of players	x	x	x	x	x	x

Micro Level: Learner and Teacher

At the micro level, both teachers and learners may benefit from all types of statistics.

Learner

The learner (possibly a player team, depending on the SG) is interested in how she has performed in a given game, where she has competence gaps and how she performed in comparison to other players or teams. A dashboard with relevant statistics is therefore useful to learners because it allows them to judge their own learning experience [16], although a teacher might be helpful in interpreting the results correctly. Following our gradation of aggregations, here are additional examples:

1. Single game play: the percentage of correct answers is already available in many SGs along with a trace of the game play, which is needed for identifying and understanding one's own errors (self-assessment) [2][20].
2. Classroom session: these data enable a detailed comparison with other players in the same class. Such statistics are often available in massive online games.
3. History of past sessions (same course): a player's history enables comparison to previous players of equivalent expertise.
4. Comparison to sessions with different categories of players: the learner can examine what specialists would have done in the same situation and—in a more differentiated way than a simple score—how far she is from an expert status in the given subject field. Active learning may thus be promoted by providing students with a comparison between their own behavior and an expert behavior.

Teacher

The teacher's concerns are twofold: how to monitor and assess students and how to improve her teaching strategies in the longer run. Research shows that well-implemented targeted instructions can significantly improve student learning outcomes [13]: trace data and resulting statistics from SGs may be very helpful in this regard.

1. Single game play statistics are helpful for computer-assisted learner assessment and identification and for coaching of students at risk (see e.g. [2]). To support this, high-level game variables must be accessible to the teacher in order to enable openness, trust and customization; in other words, the SG should not be a black box [21]. More generally, the teaching-learning cycle (e.g., as found in [5]) begins with an identification of each student's learning needs, informed by data. Such data help teachers with pacing of the provision of appropriate resources.
2. Classroom session traces will enable the teacher to monitor global progress during play and to provide additional instructions in time if the dashboard reveals problems at specific places in the SG. They may reveal whether a complementary lecture is required. After the game, statistics can also serve as a tool for adapting the assessment to the average of the class [21].
3. With detailed statistics of past sessions of the same course, the teacher can gradually develop performance standards, improve her teaching in a targeted way and measure the impact of these adjustments. Such statistics will also help in developing predictive success/failure patterns [1].
4. By comparing sessions played by different categories of learners, the teacher can better tailor her course to a given target audience. During or right after the playing,

the classroom's performance can be compared to traces previously generated by specialists; this will empower the teacher with concrete examples that she can give to the students to better illustrate which behavior was actually expected.

Meso Level: Program Manager and Game Designer

As we move away from the micro level, we suggest that stakeholders' interests will essentially focus on the wider data scopes.

Program manager

Designers and managers of educational programs want to identify subject fields to be included in a given course and preliminary competencies to be required from the participants. For postgraduate or professional training, their questions are: what are the learning outcomes and which competencies are needed to perform a specific activity, what is the specific training needed by a new employee, is the content different if the new employee has a scientific academic background or a managerial education, etc.

1. Single game play: weakly informative without other data to compare to.
2. Classroom session: if student backgrounds are not homogeneous, session statistics may pinpoint insufficient prerequisites in certain pathways.
3. History of past sessions (same curriculum): these data allow adjustment of course content and complexity to the evolution of player results.
4. Comparison by categories of learners: the definition of a training program may be based on statistical comparison of behaviors in a specific SG, with analysis of differences between a cohort of experts (the control group) and a typical learner following the training program. It might also be possible to check whether there are fewer differences with the expert control group at the end of the program than before.

Game designer

Developers and designers of SGs are interested in improving the design or contents of existing games and in developing new games that are more captivating and effective in terms of learning.

1. Single game play: weakly informative without other data to compare to.
2. Classroom session: to identify significant playability/usability issues in new games.
3. History of past sessions (same game scenario): to validate and improve game ergonomy and interest, e.g., if no player ever went through a given branch of the scenario, that branch should probably be made more attractive or else suppressed. Performance standards inside the game may also be adjusted on the basis of statistics.
4. Comparison by categories of learners: this may constitute a benchmark enabling automatic recognition of a player's profile. When the game identifies a player with little expertise in a given subject field, it may propose additional exercises.

Macro Level: Researcher and Decision Maker

Researcher

Researchers in pedagogy may be interested in the real impact of SGs and have questions such as how effective a specific SG really is and which competencies are developed inside a given game. They are also concerned with how the impact of an educational program or of different pedagogical approaches can be measured.

1. Single game play: weakly informative without other data to compare to.
2. Classroom session: comparison of learning paces among students in the same class may help devise a range of effective teaching strategies to be taken up by teachers.
3. History of past sessions (same course or curriculum): data on game play over time may help in building up a clear and comprehensive knowledge of students' needs in order to enhance pedagogical methods. On the other hand, emerging new technologies for tracking psychophysiological signals may reveal how efficient the learning process is inside an SG and how engaging a game really is [18]: the availability of historical data will contribute to establishing more reliable correlations between player behavior, learning outcomes and neurophysiological traces.
4. Comparison by categories of learners: this may help researchers understand and improve the effectiveness of pedagogical methods (including SGs) for various professional sectors, using the SG to assess the acquisition of knowledge. This may help them understand when instructional methodologies need to vary in order to accommodate differences in student learning outcomes.

Other decision makers

This is a category of decision makers who may use statistics of behaviors inside the virtual world in order to support decisions made in the real world.

1. Single game play: weakly informative without other data to compare to.
2. Classroom session: from an HR perspective, behavioral statistics can be used to help decide between candidates competing for the same position.
3. History of past sessions (same category of players):
    - HR department: to assess and improve employee competency.
    - Academic managers (faculty or institution-level): to follow global student performance through the years and increase accountability [2].
4. Comparison by categories of learners:
    - HR department: to sort employees by level of expertise and to optimize deployment of human resources.
    - Academic managers: to adjust curricula and ensure they fit the needs of the labor market.
    - Policy makers: to compare performances of schools or institutions [6].

3 A Study using the Wegas Platform

Wegas is our web-based serious game authoring and execution platform[1]. Based on several years of experience in education in diverse environments, Wegas has been designed with a strong focus on scenario definition capabilities, in order to allow educators to customize content to learning objectives and to the participants' specific subject fields. The platform supports both hybrid on-site classroom learning and remote e-learning.

Wegas serves as the basis for a broad range of games, essentially of educational nature, of which the most popular is the Project Management Game (described below), a simulation tool for budding project managers that is actively used in many universities in Switzerland and France. Wegas includes a teacher dashboard offering a real-time overview of player positions in the game, which helps identify students in difficulty. The dashboard also enables the teacher to inspect or impact player sessions, such as by giving advice inside the game as if it were coming from one of the virtual characters.

The generation of trace data is integrated in the platform, which systematically logs all choices made by the player (i.e., answers to questions or decisions made inside a game) as well as all values taken by numeric variables defined inside the game (representing e.g. game phases or player performance indexes).

3.1 The Project Management Game

The Project Management Game (PMG) was designed as a complete educational concept for teaching project management in a team-based approach that fosters collaborative skills [17]. This concept combines a simulation game with ex-cathedra theory and real-world activities such as presentations to a project management office.

In PMG, players take the role of project managers and have to choose among proposed actions. An important aspect of the game is the absence of judgment: there is no explicit feedback after each action to tell if it was a good or a bad choice. Instead, some general performance indicators will be updated from time to time and messages will arrive from virtual stakeholders inside the game (clients, colleagues, company executives, etc.). Learners therefore have to evaluate by themselves the impact of their actions and to conceptualize on their own the knowledge that could be built on this experience [14]. Following the principles of experiential learning [15][19], the game teaches them to learn and to experiment with their new knowledge inside the simulation.

3.2 The Study

Over the course of one year, trace data were collected from PMG as it was used by 5 different teachers in 10 different courses. A typical classroom comprised 5 teams of 4 students each. For the study, a new interface was developed that enables the selection

---

[1] Project home page: http://www.albasim.ch/
The platform is open source and can be downloaded from https://github.com/Heigvd/Wegas

of up to four "groups" (aggregations of training sessions) and the display of statistical comparisons between the corresponding cohorts.[2]

For example, for a decision about doing a market analysis at the very beginning of the PMG game, we observed similar behavior in three classes of students who had mainly an engineering background (bachelor in media engineering, certificate of advanced studies in energy management, diploma in project management). Less than 30% of them decided to perform a complete market analysis, and a majority chose to do only a partial market analysis in order to preserve their budget and timeline. When the same question was given to different classes of students enrolled in management curricula (bachelor in management), we observed a similarly consistent behavior, but this time with a majority of them deciding on a complete market analysis and no more than 30% percent choosing a partial analysis. We found an analogous correlation among students in business administration (MBA).

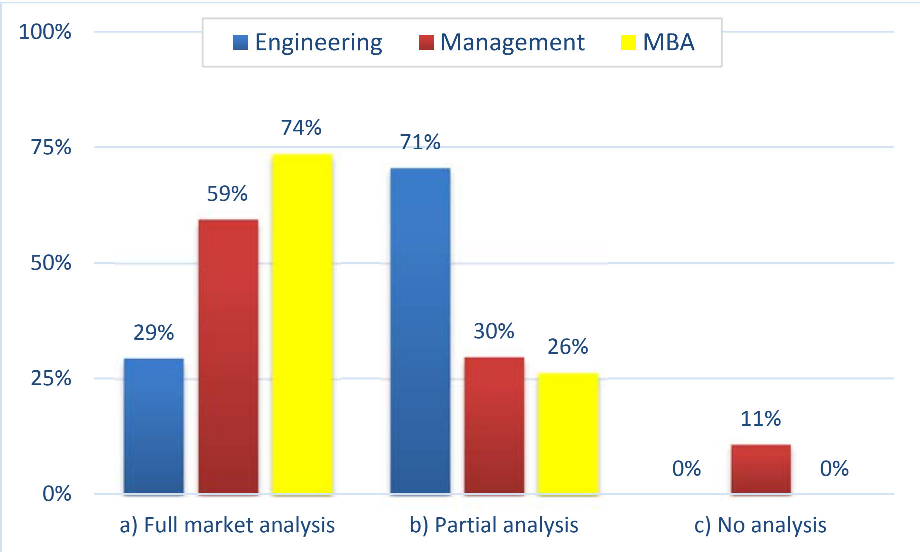

Fig. 1. Comparing decision making among three cohorts with different backgrounds.

In Fig. 1, players with similar backgrounds are grouped in order to compare typical behaviors of students with an engineering profile (blue bars), a management profile (red bars) and an MBA (yellow bars). In this graphic, one can clearly observe that students enrolled in engineering curricula are less likely to perform market analysis than students in management studies. The graphic also indicates that students in an MBA program tend to behave like students of management, even if these MBA students have an engineering background.

[2] As of this writing, some of the functionality is only available to a restricted set of users until most usability and privacy protection issues have been addressed.

4 Discussion

We have proposed a classification of stakeholders of SGs and of the data scopes that we believe might be of interest to them. We developed an extension to the Wegas platform in order to experiment with these statistics, using a significant amount of data generated during one year by different categories of players in the PMG game. We observed that the comparative analysis of behaviors may bring excellent opportunities for pushing the adoption of SGs in the learning process, improving existing games and supporting decisions based on behaviors observed in SGs.

4.1 Improving Education Management

The comparison of different groups of learners may have a positive impact on decisions in the professional world or when designing training programs. By this means, it becomes possible to answer questions like "in this specific situation, what are the differences in behavior between an engineer and someone with a managerial background?" or "what should be included in an MBA program for engineers in order to give them a managerial behavior?"

4.2 Use in Professional Assessment

Comparative statistics also enable the detection of situations in which senior professionals have different behaviors than junior professionals or bachelor students. This may impact business decisions like "what kind of qualifications are needed for this kind of activity: expert, senior or junior?" Then, one can identify players that have a behavior similar to recognized experts but are less expensive to hire. One could more easily find new positions where valuable collaborators can be relocated inside a company by fitting their player statistics into specific categories. Thus the target audience of such analytics may be widened to HR specialists of competency management, whereas the traditional audience is composed of teachers and learners [6][23].

4.3 Pedagogical Impact

With Wegas the teacher has the opportunity to present comparative statistics to the students and to highlight places in the SG where they behaved differently than actual expert-level players. Feedback based on previous cohorts constitutes a complete Learning Analytics Cycle, as recommended in [6]. Wegas makes this concept even more effective, as it enables the teacher (1) to prepare striking comparisons by selecting classrooms of different levels of expertise and (2) to provide illustrated feedback immediately after the playing session, which probably is the best moment to explain the reasoning behind an expert behavior. Emphasis may thus be placed on comparing reasoning instead of final scores, which warrants the pedagogical value of the feedback. Moreover, research shows that immediate feedback is beneficial to the learning process [8].

This way, the teacher can provide a significant added value in comparison to a situation in which statistics would simply be made available to the learner as a self-service.

4.4 Limitations and Perspectives

This article describes work in progress. Our goal was to identify and to classify useful statistics for various kinds of stakeholders, but we do not claim to be exhaustive at this stage. Whereas comparative statistics have been employed to enhance lectures using the Wegas platform and to make game scenarios more relevant, it's still necessary to quantify more generally the impact of such statistics on the learning experience. Recent research already brings evidence that the provision of immediate feedback promotes student engagement in the learning process [8].

We have currently only tested our approach on the comparison of answers to closed questions. It is important to limit the number of possible answers, because this enables more reliable and objective comparisons. In the comparative approach, we are indeed mainly concerned with how close a player's performance is to that of various control groups. In order to extend this work, we may study the chronology and speed of action sequences and try to compare them. In addition, we have not tried to take into account communication patterns between players (chats, forums, etc.). This topic, which may reveal players' level of interest or a possible need for help, is already studied in [24].

A few requirements have to be fulfilled to make statistical comparisons reliable:

- A sufficient number of sessions must be played and logged beforehand.
- Game sessions should be organized with consistent player profiles in order to enable their classification into distinct categories. On the other hand, such categories do not necessarily have to be known a priori: one interesting research direction would be to apply data mining techniques on the statistics in order to elicit novel clusterings and comparison criteria after the game is played.
- Game scenarios and content must remain sufficiently stable across versions: comparisons are currently only allowed on choices that have the same internal identifier (chosen by the scenarist). This constraint will have to be relaxed, since new game versions will necessarily be published in order to correct errors, but also to observe new aspects of players' behaviors or to encompass teachers' customization requests.

Regarding privacy protection, complete playing traces are indeed collected and processed internally, but nominative data are only accessible to the player who produced them and to the relevant teacher. No personal information is made visible by the new statistics modules, except that teachers might be identified indirectly by cross-checking. Nevertheless, data protection issues in SGs need to be further investigated [22], especially as broader data aggregations will affect a larger number of persons and might be governed by legislations of multiple institutions and countries.

Finally, usability is an important requirement: the SG platform has to make statistics easily accessible to ensure that they will actually be employed by stakeholders such as teachers and students.

5 Conclusion

In this paper we proposed a vision of learning analytics that focuses on data aggregation levels and on their usefulness to main stakeholders. A systematic classification revealed a number of useful statistical comparisons that we have not found elsewhere in the literature. We extended our Wegas platform with a new aggregation and comparison module, which was fed with a significant amount of trace data collected during one year from different categories of players of our project management game. This experience confirmed that player profiles are relatively homogeneous inside each pathway and that complementary schooling allows players to improve their behavior in comparison to their initial weaknesses. Comparative statistics enable teachers to provide striking feedback based on expert cohorts in order to reinforce the impact of serious games on students. Outside the educational setting, businesses and organizations can compare cohorts of employees in order to appoint them to optimal roles.

Acknowledgments. The authors would like to express their gratitude to all developers of the Wegas platform as well as to the teachers who gave their feedback on the tool.

References

1. Arnab, S., et al.: Towards performance prediction using in-game measures, Procs of the annual meeting of the American Educational Research Association, Chicago, USA (2015)
2. Arnold, K.E., Signals: Applying Academic Analytics, in EDUCAUSE Quarterly (2010)
3. Bloom, B. S. et al.: Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. New York: David McKay Company (1956)
4. E.A. Boyle, T.M. Connolly, T. Hainey The role of psychology in understanding the impact of computer game, Entertainment Computing, 2 (2011), pp. 69–74
5. Bybee, R.W.: Achieving Scientific Literacy: From Purposes To Practice. Portsmouth, NH, Heinemann (1997)
6. Clow, D., The Learning Analytics Cycle: Closing the loop effectively, Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12 (2012)
7. Deslauriers, L., Schelew, E., Wieman, C.: Improved Learning in a Large-Enrollment Physics Class, Science , Vol. 332, Issue 6031, pp. 862-864 (2011)
8. Dumont, A., Berthiaume, D.: La pédagogie inversée, De Boeck, Louvain-la-Neuve, Belgium (2016)
9. Entwistle, N.: Understanding classroom learning, Hodder and Stoughton, London (1987)
10. Ferguson, R.: Learning analytics: drivers, developments and challenges. International Journal of Technology Enhanced Learning, 4(5/6), 304 (2012)
11. Freeman S., et al.: Active learning increases student performance in science, engineering, and mathematics. Procs of the National Academy of Sciences, 111, pp 8410-8415 (2014)
12. GALA R&D Roadmap on Serious Games no.3, Deliverable no. 1.7 (4th year) of the Game and Learning Alliance (GaLA), 2014. http://www.galanoe.eu/index.php/documents
13. Goddard, Y., Goddard, M.: A theoretical and empirical investigation of teacher collaboration for school improvement and student achievement in public elementary schools. Teachers College Record Volume 109 (4), p. 877-896 (2007)

14. Jaccard D., Riboni S.: Simprojet: An Innovative Simulation Platform for Experiential Learning in Project Management, in Proceedings of the 2nd International Conference on Computer Supported Education, pp. 471-477, Valencia, Spain (2010)
15. Kolb, D.A.: Experiential learning: experience as the source of learning and development, Journal of Organizational Behavior, Vol. 8 (4), pp. 359-360, Prentice Hall, NJ (1984)
16. Marsh, C. J.: Key concepts for understanding curriculum, Falmer Press, London (1997)
17. Michaelsen, L., Sweet, M. & Parmelee, D.: Team-Based Learning: Small Group Learning's Next Big Step. New Directions in Teaching and Learning no. 116, (2009).
18. Moreno-Ger, P. et al., Learning Analytics for SGs, Deliverable no. 2.4 (4th year) of the Game and Learning Alliance (GaLA), 2014. http://www.galanoe.eu/index.php/documents
19. Rogers C.R., Freedom to Learn: a view of what education might become, Columbus, Ohio, C.E. Merrill Pub. Co., pp. 111-114 (1969)
20. Sanchez, E., Ney, M., Labat, J-M. : Jeux sérieux et pédagogie universitaire: de la conception à l'évaluation des apprentissages. Revue Internationale des Technologies en Pédagogie universitaire, Pédagogie Universitaire Numérique, 8 (1-2), pp. 48-57 (2011)
21. Serrano-Laguna, A., Torrente, J., Moreno-Ger, P., Fernández-Manjón, B., Tracing a Little for Big Improvements: Application of Learning Analytics and Videogames for Student Assessment, 4th Intl Conference on Games and Virtual Worlds for Serious Applications (2012)
22. Slade, S, Prinsloo, P., Learning Analytics: Ethical Issues and Dilemmas, American Behavioral Scientist, 57(10), pp. 1510-1529 (2013)
23. Verbert, K, et al., Learning Analytics Dashboard Applications, American Behavioral Scientist 57(10), pp. 1500-1509 (2013)
24. Verbert, K, et al., Learning dashboards: an overview and future research opportunities, Journal of Personal and Ubiquitous Computing, Vol. 18(6), pp. 1499-1514, Springer (2014)
25. World Economic Forum: New Vision for Education: Fostering Social and Emotional Learning through Technology (2016)