Proceedings of
The 8th European Conference on Games Based Learning
ECGBL 2014

Research and Training Center for Culture and Computer Science (FKI)
University of Applied Sciences
HTW Berlin
Germany
9-10 October 2014

Volume One

Edited by
Professor Dr.-Ing. Carsten Busch

Replacing PISA With Global Game Based Assessment

Harri Ketamo[1] and Keith Devlin[2]
[1]Satakunta University of Applied Sciences, Finland
[2]Stanford University, USA
harri.ketamo@samk.fi

Abstract: International Educational performance measures, like the Programme for International Student Assessment (PISA), rank countries based on their educational system's performance. However, these activities require years of work for thousands of people to collect, analyze and report the data. Furthermore, the results are always two years old when published and there are tens of missing countries. In this study, a global game-based analytics was introduced and evaluated in a real world environment. The paper continues the study reported in ECGBL 2013. Real-time game-based assessments do have several advantages compared to traditional measures, but they are not yet ready to replace PISA. For example, sample collection has to be radically improved in order to minimize noise and bring out only the valid cases.

Keywords: assessment, games based learning, learning analytics, mathematics learning

1. Introduction

International Educational performance measures, like the Programme for International Student Assessment (PISA), Trends in International Mathematics and Science Study (TIMSS), and Progress in International Reading Literacy Study (PIRLS) rank the countries based on their educational system's performance. They also describe socio-economic backgrounds and factors behind outcomes. Such activities are valuable in order to understand differences in learning as a global phenomenon.

However, these activities requiree years of work for thousands of people to collect, analyze and report the data. Furthermore, the results are always two years old when published and there are tens of missing countries. Finally, the biggest missing part is the understanding how to improve the system at a national or individual level.

Educational games and game-based assessment can be seen partial answer to this: distribution of the assessment (i.e. games) can be done globally without any extra effort, using outlets such as the Appstore. The data is always in real time, and if the instruments are designed in terms of learning science, they will measure learning and learning outcomes. Arguably the most important advance is the capability to give suggestions how to improve curriculum, teaching methodology, or learning materials. Furthermore, real-time analytics enables individual-level recommendations and adaptation in learning.

In this paper we introduce one approach for real time game-based assessment, show global assessment results based on game data, and discuss the future direction of research and challenges to be overcome before game-based assessment can replace PISA or TIMSS. The paper continues the study reported in proceedings of ECGBL 2013 (Ketamo 2013).

Behavior modeling has a long background in educational technology: Neural and semantic networks, as well as genetic algorithms, are utilized to model users' characteristics, profiles, and patterns of behavior in order to support or challenge the performance of individuals. Behavior recording has been studied and used in the game industry for a long time. In the majority of studies, the level of behavior is in general limited to observed patterns (e.g. Brusilovsky 2001; Houllette 2003; Bowling & al 2006; Bra & al 2013; Shyi-Ming & Po-Jui 2013).

In this study, user behavior, competence, and learning are seen as semantic (neural) network that is inductively taught by the user. According to the cognitive psychology of learning, our thinking is based on conceptual representations of our experiences and relations between these concepts. Learning is what occurs when the mental structure changes. Data mining and analytics are based on this semantic model. When all the skills and knowledge are recorded as a semantic network, the mining can be done in terms of network analysis. The novelty of this study lies in the approach: to build game-based technologies that enable easy inductive construction of intelligent and human-like behaviors, and so enable detailed analysis of learning achievements. The same kind of competence modeling, based on deductive learning theories, has been done by e.g. Biswas, Leelawong, Schwartz & Vye (2005).

We know that children are ready to do more work for their game characters than for themselves (e.g. Chase & al. 2009). That is why Skillpixels (a company Ketamo has founded in 2011) has developed games where game characters learn like humans do, with children taking the role of a teacher. The framework and technology behind the games support detailed learning analytics and provides real time analysis of learning processes, difficulties in learning, and challenges in the curriculum.

2. Research procedure

The sample (N=49080) was collected with SmartKid Maths Online version, between 1.8. and 31.12.2013. This online version was targeted especially for school piloting, which helps us to make an assumption that most of the players were 1st and 2nd grade pupils. Only the players who had completed at least 10% of the content (approximately 1.5 - 2 hours of game play on average) were included into this sample. This decision was made to ensure that all the included data is collected from pupils who really played the game, not just tried it briefly or had a short period of use.

Because all collected data is completely anonymous, we can be sure only about the country that the data comes from. We cannot guarantee that all the first grade players are really first grade pupils, and so on. Furthermore, the nature of the game is that when a player has completed kindergarten, he/she continues to first grade. This means there are a lot of kindergarten pupils included in first grade data and a lot of first graders in second grade data. In this paper, we focus on describing the procedure and findings, so the sampling is not a problem, but when describing learning achievements and competences in details, this is maybe one of the most critical challenges to counter.

In this paper, only countries with more than 1500 players were included in the analysis. The final sample, 1000 pupils per country, was randomly selected from all the data from the specific country. The differences in the results between three different random samples were below 2%.

Furthermore, the sample was divided in two subgroups: casual players and hard gamers. Casual players were those who completed more than 10% but less than 35% of the game. In other words those who played 2-6 hours on average with several sessions. The hard gamers group consist of those completed more than 35% of the content. Some of them played only 5 hours with extremely good performance, and some players spent more than 100 hours with SmartKid Maths, ending up with a good performance.

However, most of the pupils who played more than 35% of the content were kindergarten pupils who played first grade content or first grade pupils who played second grade content. When analyzing the results, we noted that the more levels player opened, that more he/she faced completely new topics.

In the results section we provide analysis from Finland, Sweden, United States, Canada, United Kingdom, Ireland, Malaysia, India, Latin America (as a group in order to reach the defined N), and Australia.

3. Research materials

SmartKid Maths is a game-based, virtual mathematics school for mobile, tablet and online users (Figure 1). The biggest difference from other mathematics games is that in SmartKid Maths the player teaches the computer, not vice versa. When the player is responsible for character's mental development, he/she records also his/her mental conceptual structure during the gameplay.

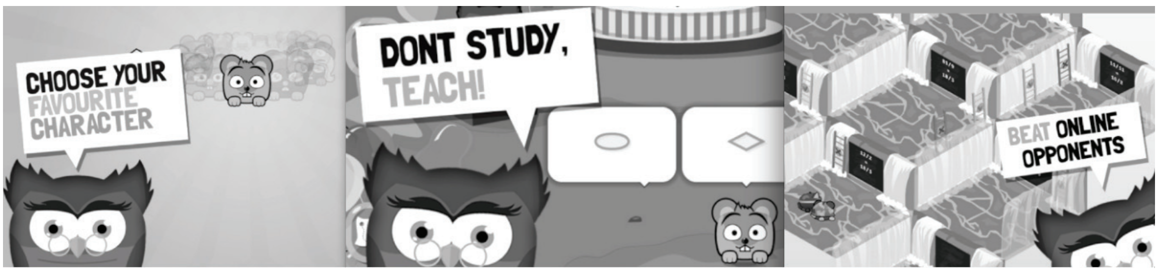


Figure 1: Collection of screenshots from SmartKid Maths

SmartKid Maths include all mathematics in the US common core K-2 curriculum, which represents approximately 300-400 pages of exercises in traditional books. In the game, there are 100 levels and one level represents approximately one school week, 3-5 pages of the exercise book, in traditional school. SmartKid Maths covers the following Common Core topics:

Grade K: Counting and Cardinality

- Know number names and the count sequence.
- Count to tell the number of objects.
- Compare numbers.

Grade 1: Operations and Algebraic Thinking

- Understand and apply properties of operations and the relationship between addition and subtraction.
- Add and subtract within 20.
- Work with addition and subtraction equations.

Grade 2: Operations and Algebraic Thinking

- Represent and solve problems involving addition and subtraction.
- Represent and solve problems involving multiplication and division.
- Add and subtract within 100.
- Multiply and divide within 100.
- Work with equal groups of objects to gain foundations for multiplication.

One of our special foci has been the scientific proof of concept: We have shown educational outcomes as well as motivation towards teaching virtual pets: Under strict laboratory experiment settings, more than 60% of players increased their skills remarkably during the gameplay. The outcome in a natural learning environment, with the possibility of longer gameplay, is even greater: the best outcome is achieved when there are breaks and informal discussions between gameplay.

When the player is responsible for their character's mental development, he/she records also his/her mental conceptual structure during the gameplay. The most important finding is that assessment carried out according to learning data collected during the game play correlates with assessment done with traditional paper tests: Taught conceptual structure is strongly related to paper tests scores received after gameplay ($0.4<r<0.7$) with all tested content on mathematics and natural sciences. This is an important result in terms of the reliability of the game as an assessment/evaluation instrument. Because of this, we can produce detailed diagnostic information about learning (Ketamo 2009, Ketamo & Kiili 2010, Ketamo 2011).

In Skillpixels' Math Elements game, every character is an agent/entity that contains all the taught knowledge as a semantic network. That semantic network can be used to produce character behaviors, including reasoning, without connection to the original game or other characters (where the knowledge is recorded). Because every agent is taught by an individual person, each agent has different skills. Furthermore, as noted earlier, the knowledge and behavior recorded in an agent's semantic network correlates with players' real world knowledge. Overall, the agents / game characters are ideal entities to run different simulations, which can be presented as competitions as well as complex data mining procedures.

Teaching the character can be approached in terms of inductive learning theories. The general idea behind inductive learning theories is that we build our understanding by processing and connecting single concepts into large conceptual understanding piece by piece. Adding new concepts or modifying the existing conceptual structure is always based on previous learning and the context of learning. Because of that, both learning and mental conceptual structures are unique for everyone.

Inductive learning has been applied in Math Elements in a way where the player teaches the character item-by-item in different learning contexts. While playing, the conceptual structure will grow to thousands of relations and a single teaching phase only has a limited effect on the areas of the conceptual structure already taught. Understanding this phenomenon is valuable when trying to correct a wrongly taught part of the conceptual structure. Naturally, wrong teaching could be corrected by teaching the correct structure enough times. The game AI uses all of the taught information to support its decisions, and therefore it takes time to override any wrong relations in the agent's conceptual structure.

4. Results

Analytics outcomes of the learning process can be used when developing national curricula or learning materials. When summarizing the individual game achievements (taught conceptual networks), schools and national level policy makers can receive analysis about learning achievements and competences at a general level. They can apply this to develop their teaching instructions or formal curriculum. Our goal in this paper is not to rank countries, but to provide information that might be useful for developing curriculum or classroom practices.

In this analysis we compare differences in understanding and applying the core concepts of numeracy:

- count of objects
- number symbol
- place in number line

The game records all the events during the gameplay. First we identify the number of difficulties or recorded misconceptions encountered during the game play. After that, identify players who have overcome their previous difficulties when the level or game is completed.

On the left side of Tables 1-6 are the percentage shares of pupils who mastered the content all the time during the game play (below 5% misconceptions at any stage of the gameplay) and on the right side the percentage shares of pupils who mastered the content after the level or game is completed. The difference between these shares shows the effect of the game: the more pupils masters the content after the game play, that better the game performs as a learning game. In Tables 1-3, we focus on casual players group (completed more than 10% but less than 35% of the game). Focusing only on a conceptual level, we can see that the concept of number symbol is the easiest one. In other words, pupils recognize the written numbers and only a relatively small number of pupils had problems applying number symbols. The number line was the most difficult concept. Surprisingly, applying counts was more challenging than the number symbols. This phenomena was observed in all the countries.

Table 1: Mastering the concept of Number (as symbol) during and after the gameplay. Players who played more than 10% but less than 35% of content are included in figure

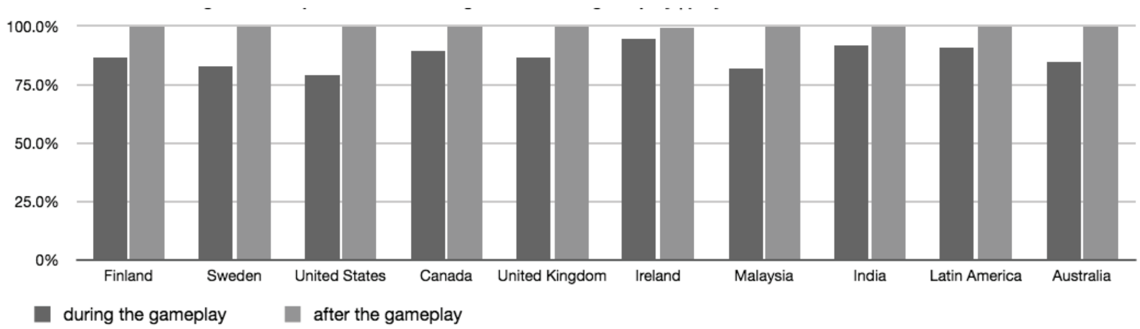


Table 2: Mastering the concept of Count during and after the gameplay. Players who played more than 10% but less than 35% of content are included in figure

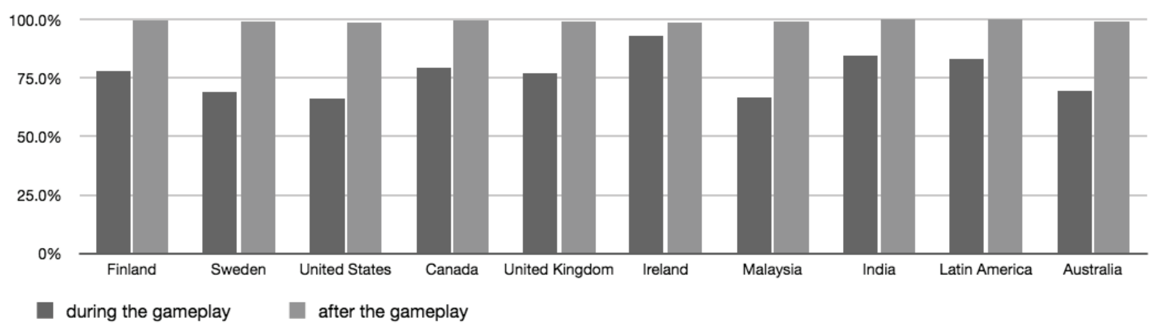
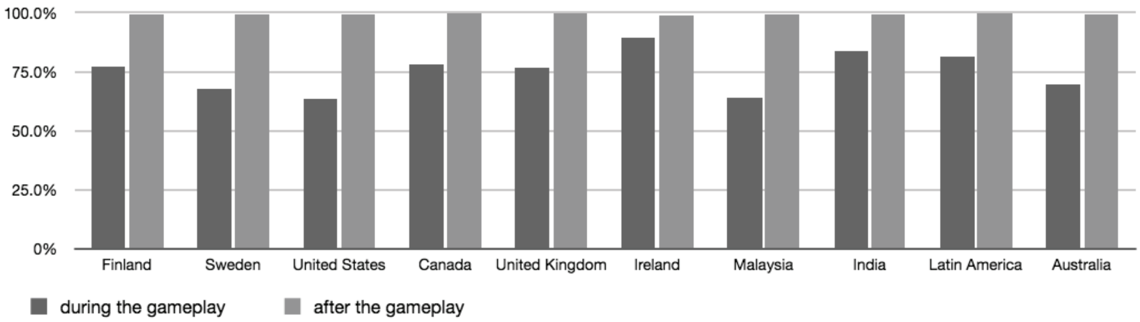


Table 3: Mastering the concept of Number Line during and after the gameplay. Players who played more than 10% but less than 35% of content are included in figure



When comparing countries, there are relatively big differences in percentage shares of those mastering the content. We would expect to find statistically significant differences, but because our data collection and sample definitions were relatively superficial, we did not run any statistical analysis on this sample. Instead, we chose to describe the method and report its strengths and weaknesses.

Most of the differences can be explained by following two facts: 1) Kindergarten is started between 3 and 6 yeas of age, depending on country. This kind of difference in age obviously affects to performance in the game. 2) We know many kindergarten pupils played the first grade game, but we also know that in some countries we had remarkably big sample from first and second grade. When doing a random sample, the big share of second graders affects the sample. In Tables 4-6 we focus on the hard gamers group (completed more than 35% of the content). This group is very different from the casual players group: Most of the players completed content (levels) that goes far beyond their school literacy. In other words, kindergarten pupil who have played half of the first grade content belongs into this group. Furthermore, is seems that those pupils who faced challenges during the gameplay and overcame those challenges experienced kind of a Flow experience,and therefore played for longer than those who did not learn as much during the gameplay. From Tables 4-6 we see that this time the concept of count was the easiest one. This is mostly because of the tasks in the game: The concept of count was excluded from the most difficult tasks and hence the pupils did not face that big challenge with counts.

Table 4: Mastering the concept of Number during and after the gameplay. Players who played more than 35% of content are included in figure

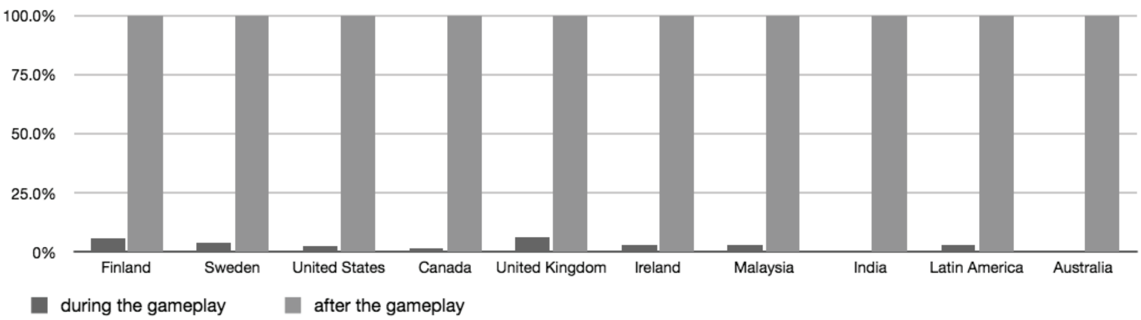


Table 5: Mastering the concept of Count during and after the gameplay. Players who played more than 35% of content are included in figure

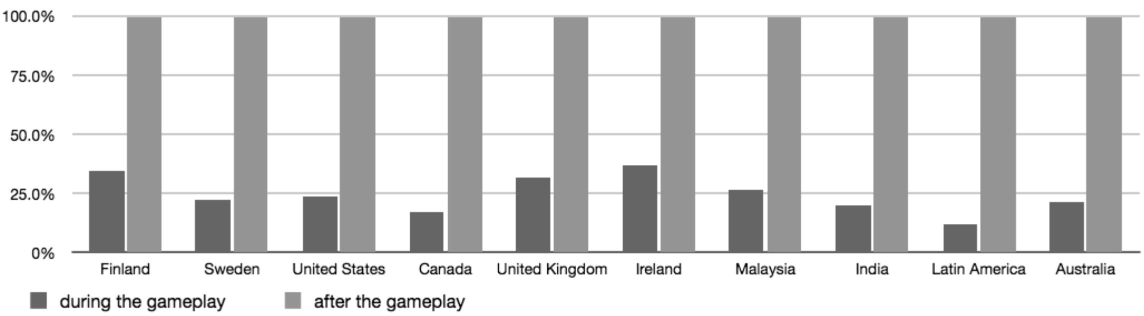
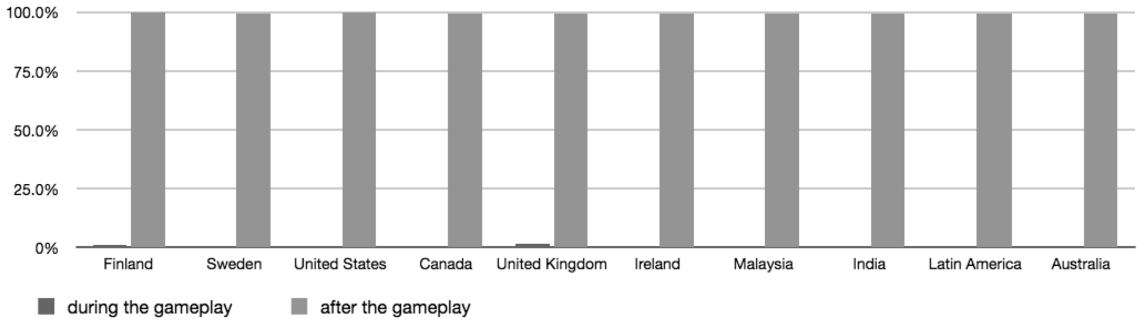


Table 6: Mastering the concept of Number Line during and after the gameplay. Players who played more than 35% of content are included in figure



Learning is about changes in our understanding. Educational games in general are meant to provide good learning outcomes. This has been the most important goal for SmartKid Maths. When focusing on the right side of Tables 1-6 we can see that almost all the of pupils mastered the content after the gameplay, no matter how many did not have mastery before the gameplay.

This in not only because of the method and content. It is also because of game-type storytelling: When paying games, children are used to working hard in order to access to the next level. Before a player can access the next level, he/she must have enough skill from previous levels. The same is not the case for school, where pupils can continue with the next topic no matter how many basic skills they are missing.

Because SmartKid Maths is designed to provide adaptive and personalized learning, it ensures all the basics are learned before pupil are allowed to continue. In other word, when analyzing the latest performance, almost all pupils had achieved the basic numeracy skills during the game play. This gives a strong background to continue addition, subtraction, multiplication, and division.

5. Conclusions

According to our studies, users can relatively quickly and easily teach their game character. The behavior modeling makes it possible to emulate conceptual learning and thus uncover the frequencies, dependencies, and patterns behind conceptual change and learning transfer. From the data we can determine the performance indicators for conceptual level and can compare the performance between countries. In general, the results in this study differs from e.g. PISA results, which is both an expected and an interesting observation.

Real-time global game-based assessment, described in this study, is not yet ready to replace PISA 2015. We have to radically improve our data collection and sample design in order to bring out really valid results. This might be done via schools: when schools can anonymously identify the population, we'll be able to eliminate unknown players and filter out conceptual changes resulting from long playing time.

One future research activity to decrease this sampling challenge could be based on estimating when pupils play at a level that is completely out of their competence. We have observed that pupils who have played through a lot of content had major conceptual misunderstandings during the gameplay, while those who played less did not demonstrate such misunderstandings. If the onset of critical conceptual misunderstanding could be identified, we could separate first graders from second graders and kindergarten pupils from first graders.

In general, games provide a different view of learning because of the time frame. Learning is about change and change requires time. When we can provide valid analysis about learning process and learning assessment, we can provide something new in addition to PISA, TIMSS and PIRLS.

References

Biswas, G., Leelawong, K., Schwartz, D., Vye, N. (2005). Learning by Teaching: A New Agent Paradigm for Educational Software. Applied Artificial Intelligence, 19 (3-4), 363–392.

Bowling, M., Furnkranz, J., Graepel, T. & Musick, R. (2006). Machine learning and games. Machine Learning, vol 63, pp. 211-215.

Bra, P., Smits, D., Sluijs, K., Cristea, A., Foss, J., Glahn, C. & Steiner, C.M. (2013). GRAPPLE: Learning Management Systems Meet Adaptive Learning Environments. In Peña-Ayala, A. (ed.) Intelligent and Adaptive Educational-Learning Systems: Smart Innovation, Systems and Technologies. Volume 17, 2013, pp 133-160.

Brusilovsky, P. (2001). Adaptive Hypermedia. User Modeling and User- Adapted Interaction, vol 11, p. 87-110.

Chase, C, Chin, D, Oppezzo, M, Schwartz, DL. (2009). Teachable agents and the protege effect: Increasing the effort towards learning. Journal of Science Education and Technology, 18, 334-352.

Houlette, R. (2003) Player Modeling for Adaptive Games. In Rabin, S. (ed.) AI Game Programming Wisdom II. Massachusetts: Charles River Media, Inc.

Ketamo, H. (2013). Learning Analytics with Games Based Learning. In proceedings of the 7th European Conference on Games Based Learning. 3-4 October, Porto, Portugal, pp. 284-289.

Ketamo, H. (2011). Sharing Behaviors in Games and Social Media. International Journal of Applied Mathematics and Informatics, vol. 5(1), pp. 224-232.

Ketamo, H. & Kiili, K. (2010). Conceptual change takes time: Game based learning cannot be only supplementary amusement. Journal of Educational Multimedia and Hypermedia, vol. 19(4), pp. 399-419.

Ketamo, H. (2009). Semantic networks -based teachable agents in an educational game. Transactions on Computers, vol 8(4), pp. 641-650.

Shyi-Ming, C. & Po-Jui, S. (2013). Constructing concept maps for adaptive learning systems based on data mining techniques. Expert Systems with Applications. Volume 40 (7), pp. 2746–2755.