

Serious Gaming Analytics: What Students' Log Files Tell Us about Gaming and Learning

^{1*}Wim Westera, ²Rob Nadolski, ³Hans Hummel

^{1,2,3}Welten Institute, Research Center for Learning, Teaching and Technology, Faculty of Psychology and Educational Sciences, Open University of the Netherlands, Valkenburgerweg 177, 6419 AT Heerlen, The Netherlands.

¹wim.westera@ou.nl, ²rob.nadolski@ou.nl, ³hanshummel@ou.nl

Abstract

In this paper we explore existing log files of the VIBOA environmental policy game. Our aim is to identify relevant player behaviours and performance patterns. The VIBOA game is a 50 hours master level serious game that supports inquiry-based learning: students adopt the role of an environmental consultant in the (fictitious) consultancy agency VIBOA, and have to deal with complex, multi-faceted environmental problems in an academic and methodologically sound way. A sample of 118 master students played the game. We used learning analytics to extract relevant data from the logging and find meaningful patterns and relationships. We observed substantial behavioural variability across students. Correlation analysis suggest a behavioural trade that reflects the rate of "switching" between different game objects or activities. We were able to establish a model that uses switching indicators as predictors for the efficiency of learning. Also we found slight evidence that students who display increased switching behaviours need more time to complete the games. We conclude the paper by critically evaluating our findings, making explicit the limitations of our study and making suggestions for future research that links together learning analytics and serious gaming.

Keywords: Serious gaming, learning analytics, logging, behaviour, video

1. Introduction

Serious games are among the most challenging, most dynamic and most interactive learning environments. They commonly offer learners rich and interactive content, large degrees of control, freedom of movement and responsibility for the actions undertaken. Serious games are outstanding examples of adaptive systems, as they continuously adjust their responses to the learners' actions for preserving favourable conditions for playing and learning. Inherently, playing a serious game produces highly individualised data trails that reflect the player's personal choices, behaviours and performances. Now that data mining and data analytics are gaining attention among educational researchers and practitioners, serious games would be an excellent target. The Society for Learning Analytics Research, (<http://www.solaresearch.org/>) uses the topical term "learning analytics" to indicate "...the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs". This emerging field arouses high hopes for gaining new insights into educational practices and devising new ways to improve teaching and learning. These expectations certainly apply to serious games. Indeed most serious games dynamically capture user data for evaluating appropriate system responses to the player's actions. In most cases, however, progression in a game is guided by simple performance criteria: the only relevant thing would then be to check whether the player achieves sufficient performance milestones within the constraints of the game rules. Both the lack of established methods and tools for linking logging data directly to game play and practical constraints such as restrictions to time and budget, may hinder game developers to exploit the player's full history for creating detailed user model.



From an educational perspective, such focus on player performance is not necessarily beneficial for learning. Various authors [1][2] explain the difference between a performance orientation and a learning orientation: while game play tends to focus on performance, which is linked with an attitude of achieving milestones and score (in many cases under time constraints), learning requires opportunities for reflection, repetition, self-evaluation, pauses, and even the preparedness to make mistakes. Hence, the process of gaming may readily counteract the process of learning. Having completed a serious game successfully with a high score doesn't necessarily imply successful learning. This discrepancy between learning and performance will be larger as games offer more freedom of movement to the learners. For example, in well-structured drill-and-practice games such as math games or spelling games, the learning gains are likely to coincide with performance gains. However, in games that offer more freedom of movement and autonomy as associated with contextualized problem solving, adventure games, inquiry-based learning competence learning, self-directed learning, self-regulation and a wide range of 21st century skills relevant for today's knowledge workers [3][4], the quality of learning is likely to diverge from the quality of performance. So, the painful paradox is that as the learning in a game becomes more complex and the behavioural variability across individuals gets higher, the less information we have about the individual's process of learning and its effectiveness. It seems we're treating the serious game as a blackbox, which is supposed to produce predefined learning outcomes, but we're not able to verify the correctness of our assumptions and to assess the process of learning and its quality. For preserving the efficiency of learning it is highly relevant to gain insights in the individual behaviours, activities and efforts that the players exhibit in order to reach the game's performance milestones: e.g. did a player achieve the milestones in an efficient and well-considered way, or was it a thoughtless trial and error style that took a lot of time without achieving any learning gains? We have to open the game's blackbox and reveal the actual in-game behaviours of learners and analyse how these behaviours relate to learning achievements. Game logging data are a treasury of information, which are available for distilling more details about the players' learning achievements based on their wanderings and trajectories through the network of game state nodes. Exploiting the player's full history could be beneficial for analysing the process of learning, for tracing bottlenecks in game play, and for building detailed user models that are required for personalised learning and learner support.

This paper is an elaboration and extension of previous work reported in [5]. It provides an exploratory study of existing log files of serious games that are a regular and mandatory part of the master programme of environmental sciences at Utrecht University. For practical reasons the paper is constrained to a retrospect study: the logging files date back to 2008-2011. The size of the sample is 118 students. Our aim is to explore to what extent the logging data of these particular inquiry-based games are helpful to reveal meaningful behavioural patterns, variables and relationships. Complementary to the logging data, we were able to retrieve the final marks that participants obtained for their work in the games. This would offer opportunities for linking observed behavioural patterns with the effectiveness of learning. The exploratory nature of the study requires a restriction to hypothesis generation rather than hypothesis testing: for reasons of methodological soundness we cannot use the same dataset for both purposes. Our research questions (RQ) are specified as follows:

- RQ1 To what extent can we identify different gaming behaviours?
- RQ2 To what extent can behavioural characteristics be predictors of final scores?
- RQ3 Do we observe relationships between prior knowledge and gaming behaviours and performances?
- RQ4 How do students treat the video resources in the games?

In the following, we will first describe the state of the art in the emerging research field of learning analytics and connect this with new developments in analysing player data in serious games. Then we will describe the VIBOA serious games in environmental sciences and their context of use. After explaining our methods, we will present our findings and critically evaluate these.

2. The emerging field of learning analytics

Analysing serious game log data clearly fits within the wider framework of learning analytics: using the ever-growing amounts of data about learners' activities and interests for improving learning outcomes [6]. A related term is educational data mining [7][8]. While the focus of educational data mining is on methods for extracting the data, learning analytics concerns the development and application of predictive models in instructional systems [9]. However, Zouaq, Jocsimović and Gasević [10] concluded after text mining of a large number of research papers, that educational data mining and learning analytics are very much used as synonyms. Martin and Sherin [11] explain that educational researchers have been using sophisticated educational mining or learning analytics methods for decades. Especially research on intelligent tutoring systems in the early nineties collected detailed user data and analysed these for making improvements. Current work on learning analytics is clearly based on a longer tradition. In recent years datasets have grown larger and have become more easily accessible. From Learning Management Systems such as Moodle and Blackboard student logging data are easily extracted and combined with user profile data, access statistics and test scores. An early example of successful educational data mining at an institutional level is provided by the Signals project at Purdue University, which showed how student data can be used in predictive models and lead to higher grades and retention rates than were observed in control groups [12][13]. Also, growing interest in open educational content, open standards and MOOCs [14] produces big sets of learner data. Likewise, mobile learning adds significantly to the multitude of user trails.

To date a wide variety of methods and tools for the analysis of student data are available, including social network analysis, content analysis, discourse analysis, factor analysis, regression analysis, filtering and data visualisation. Recent studies in the UK [15][16] notice considerable fragmentation of initiatives, though. Within higher and further education institutes different departments seem to work independently on their own solutions, for their own purposes, with their own tools (e.g. Excel, SPSS), while using their own datasets, e.g. library data, virtual learning environment, human resources, web statistics, student records, register of attendance, sensor data, curriculum data. Research in the field mostly presents local cases and tailored solutions that aren't necessarily generalizable or transferable to other contexts. Gradually, however, harmonising efforts are made that propose generic frameworks or a shared set of approaches or technical standards. Hung, Hsu and Rice [17] view learning analytics as an extension of questionnaire-based course evaluations. They provide a hybrid framework for programme evaluation, which combines student learning logs, demographics data and end-of course evaluation surveys. The approach, which is grounded on factor analysis and decision tree analysis, is claimed to enable in-depth educational programme evaluation and to provide predictive models of course satisfaction, instructor satisfaction and final grades. Greller and Drachler [18] pragmatically identified six critical dimensions that are assumed to describe the process of learning analytics: objectives, data, stakeholders, instruments, internal limitations and external constraints. These dimensions are grounded on the textual analysis of literature abstracts and online discussions about learning analytics. The framework is positioned as a starting point of a learning analytics ontology. Alternative models tend to be more oriented on workflow and the process of analysis [19][20]. In their Educause White Paper Campbell and Oblinger [21] embed learning analytics in a quality assurance cycle that is composed of five consecutive steps: capture, report, predict, act and refine. Romero and Ventura [8] notice that tools for current datamining are complex and designed for power rather than usability. Dyckhoff, Zielke, Bültmann, Chatti, and Schroeder [22] view this complexity as a severe barrier for adoption and claim that teachers should have direct access to simple but effective learning analytics tools that should be integrated in the Virtual Learning Environment. They have developed a learning analytics toolbox (eLAT: exploratory Learning Analytics Toolbox) that can be used by teachers for being informed about the effectiveness of their courses and that allow them to explore and correlate key data themselves. Tools like these hold the promise that learning analytics will not only be a tool for research and quality assurance, but also a means of feedback, reflection and professionalisation for teachers.

Although learning analytics is generally qualified as an opportunity for improving the quality and effectiveness of learning, important concerns are raised because analytics could severely disempower and demotivate learners when they are provided with continuous feedback about their knowledge and performance gaps as compared with other students [23]. Also the capturing of unstructured personal traces across different platforms, social networks and contexts goes with some principle barriers linked with privacy protection and other legal issues [6]. These drawbacks also hold for serious gaming and should be taken into account.



3. Exploiting user data in serious games

There are two different types of logging analysis in serious games. First, in-game (or within-game) logging analysis refers to using the player's individual history for enhanced personalised interaction. Second, posterior logging analysis concerns the collection of user data at population level for the purpose of game evaluation and improvement. Figure 1 sketches the framing of the two types of analysis.

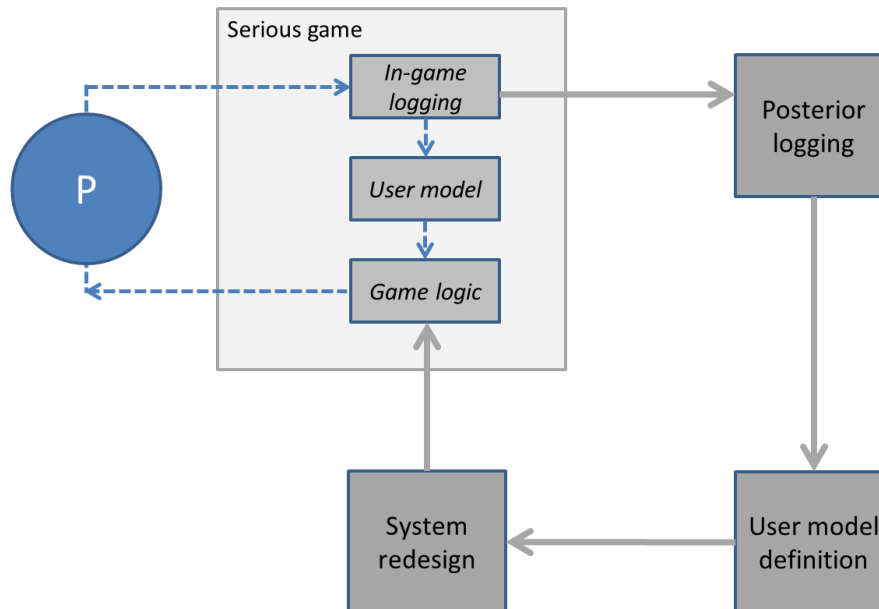


Figure 1. Framing of in-game logging and posterior logging

During game play, dynamic information from recorded actions, preferences, achievements or failures of the player (P) are logged in a personal history file and is used to extract a model of the user. The game's rule engine (game logic) uses these user data for continually evaluating the most appropriate, that is, adaptive and personalised responses the game should provide. In contrast, the posterior logging analysis (which applies to this paper) is an offline operation after game completion. It is part of an external process cycle at population level. Logged history data of multiple users are aggregated and used for quality assessment and game improvements. Relevant behavioural patterns are extracted, which are used for enhancing the user model definition and integrating this in a redesign of the gaming system. These two approaches may in turn be part of a transcending learning analytics system at curricular or institutional level.

To date, posterior logging in online games is frequently used by social gaming companies for assessing players preferences, for tracing bottlenecks in game play and for predicting what users want and will do next in the game [9]. Serious games researchers use posterior logging for achieving a better match between gaming and pedagogy. Bluemink, Hämäläinen, Manninen and Järvelä [24] report about the analysis of captured data from a multiplayer, voice-enhanced video game, aimed at revealing discourse patterns and the development of collaboration and group cohesion. The posterior learning analytics involved manual coding of recorded communication. Fernández-Gallego, Lama, Vidal, and Mucientes [25] advocate the use of learning analytics in 3D educational virtual worlds: similar to serious games, the immersive and open nature of these kind of environments offer the students a lot of moving space en encourage them to arrange their own learning. The authors present a generic learning analytics framework for 3D educational virtual worlds, which is based on 1) the IMS LD specification [26] covering the process models that describe the learning activities and 2) a script-based registration mechanism of the events and interactions generated by the students' avatars. Logged events are supposed to be fed into a set of process mining algorithms that detect and recognise predefined learning design components, or - on occasion - unwanted deviations from these. Practical application of this framework is presented as future research. Gobert, Sao Pedro, Raziuddin and Baker [27] provide a mining method for the assessment of the students' scientific inquiry skills in an online microworld. As a first step, student log data were manually categorised as to calibrate and validate a second step of automated assessment. Although the microworlds used are more like structured tutorials than serious games,

the approach may be relevant for the domain of serious games, because of the real time use of logging data during a learning activity and its potential for realtime adaptive scaffolding.

Martin and Sherin [11] discuss the potential of in-game learning analytics for improving the learning of students in an online math game, be it just as a hypothetical example. Some existing games provide within-game learner support services based on playing behaviour, e.g. scaffolding, hinting, micro-feedback, meta-level feedback [28][29]. In the case of inquiry-based serious games Westera, Nadolski, Hummel and Wopereis [4] point at using in-game logging data for strategic feedback rather than tactic or micro-level feedback, in order to nourish the learners' self-directedness, self-evaluation and reflection. Reese, Seward, Tabachnick, Hitt, Harrison and McFarland [30] report about the CyGaMEs project, which quantifies game play activity for tracking timed progress toward the game's goal and uses this progression as a measure of player learning. Serrano-Laguna, Torrente, Moreno-Ger and Fernández-Manjón [31] propose to use learning analytics for the semi-automatic assessment of learning achievements in serious games. They claim that even a small sample of learner data produces a substantial improvement of the assessment quality. Moreover, such method is an unobtrusive alternative for intermediate tests or questionnaires, which are often perceived as unwanted interruptions of game play. Similarly Shute, Ventura, Bauer and Zapata-Rivera [32] present an unobtrusive assessment methodology that heavily relies on logging data. It combines Evidence-Centered Assessment [33] and Bayesian score models [34]. Its ingredients include a competency model, a learner model, an evidence model (providing clues for evidence) and a task model. In all cases games cannot do without in-game player data, because these are straightforward requirements for evaluating the game's responses to a player's actions. The player's inputs are processed according to the game's logic, which allows for tailored responses by changing the state of the game world, e.g. adjusting the difficulty level, providing rewards for successful performances or presenting new challenges. Nevertheless, in most serious games the use of in-game logging is quite shallow: user data are mainly used for triggering events and new episodes in the game flow and game narrative, but are seldom used for the accommodation of user modelling, personalised learning and establishing learning progress [31][5]. Although player tracking is a predominant and well-exploited mechanism for adaptive gameplay, the player's full history of states is still greatly underused, because most games reflect a discrete time Markov chain, which assigns only a limited role to state history and process memory [5].

4. The VIBOA environmental policy games

The VIBOA environmental policy games of this study are a set of five online serious games that were jointly developed by the Open University of the Netherlands, Utrecht University, and Radboud University Nijmegen. These are complex and inquiry-based games that are used in the Master Degree Programmes in Environmental Sciences. The main purpose of the games is to bridge the gap between theory and practice and to provide an authentic learning context for this. The learning objectives include the methodologies of analysis, evaluation and design of environmental policies. The five games are linked together in a single run that requires about 50 hours of study load. Each game presents an authentic, multifaceted problem case that should be dealt with by the students in an academic and methodologically sound way: 1. Wadden Sea, 2. Wind energy, 3. Lake Naarden, 4. Micro pollution, 5. River management. In the games, students adopt the role of an environmental consultant in the (fictitious) consultancy agency VIBOA. Students have to play the games individually. They have to apply scientific methodologies and theories in a context that is imbued with conflicting views, conflicting interests and conflicting demands. They have to make a thorough analysis of the problems and devise solutions for these by collecting and combining relevant information from reports, scientific papers, interviews, texts of law, formal documents and other sources. Occasionally, incoming notifications or (pseudo) email messages provide new information, announce new events, provide hints or prompt for certain actions. Videoclips are an important element of the game: they allow players to have video-based meetings and interviews with experts and stakeholders, and thereby contribute to enhanced realism and sense of urgency (cf. figure 2).



Figure 2. Screenshot of the VIBOA environmental consultancy game

The first VIBOA game is an introductory game, which is concluded with a test for checking and aligning prior knowledge. For each of the other four games the students have to deliver a report, which specifies their approach, their analysis and their proposed solutions. These reports are the basis of the formal examination.

The VIBOA games are implemented in the EMERGO game engine (www.emergo.cc), which is an open source educational gaming platform developed by the Open University of the Netherlands. It allows for web-based authoring and delivery, and it captures a full set of logging data that are worthwhile analysing.

5. Method and conditions

5.1 Data collection

The EMERGO game engine entails a component-based architecture [35]. Each component allows for the authoring and structuring of a specific set of gaming objects and their attributes. EMERGO offers up to twenty different component types, for example a floor plan component is used to define the locations that make up the gaming environment, a video-interview component allows for the specification of video-based dialogues, a chat component enables chat between players, a resources components supports the arrangement of documents and other knowledge resources, a testing component supports item and test construction, etcetera. These components are linked together by the game logic, which allows for the scripting of production rules. Because of the independent agency of the EMERGO components, the logging data aren't stored in a single file but are distributed over the components. To address this issue we have implemented a logging aggregator, which collects and combines the logging data into a joint single file. It captures all meaningful student actions such as visiting a location, opening an information resource (document, URL, picture, graph, etc.), accessing a video, asking a video-interview question, accessing a pre-test item and some more. Also system responses are recorded. All actions go with a timestamp.

5.2. The data

We used data from student cohorts of Utrecht University. The games were played as part of the regular master curriculum. We have combined the data from cohorts of 2008, 2009, 2010, and 2011 in order to enhance statistical power. Across these years the same internal and external conditions were maintained (e.g. preparation, time frames, intermediate assessments, examinations). From the sample we have excluded students who didn't complete all games and

failed to obtain a final score (e.g. dropouts). The cleaned sample consisted of 118 students (46% male, 54% female) including 7 people who failed in one year but re-enrolled in the subsequent year. The resulting data file had about 700,000 records.

In addition to these logging data we were able to retrieve the examination scores that were assigned by the examiners to the students' reports about the respective games. This would in principle enable us to link observed gaming behaviours with the quality of learning. After linking the logging data and the examination marks we anonymised the dataset for further processing.

5.2. Data processing

We have confined ourselves to extracting a set of basic variables that were retrievable for each student with simple queries. Tabel 1 lists these primary variables, which will be explained below.

Table 1. Primary variables extracted from the sample

Notation	Description
T	Total time: the total time spent by a student to the 5 games
Nu	Number of user actions: the total number of the student's functional choices and decisions
NL	Number of locations: the total number of locations (re)accessed by a student
NR	Number of resources: the total number of resources (e.g. documents) retrieved by a student
NV	Number of videos: the total number of embedded video clips (re)accessed by a student
TV	Time spent to videos: the total time spent by a student to watching videos
TVL	Length of videos: the total length of the videos retrieved by a student
SP	Score on pre-test: the score of a student obtained from the built-in pre-test (initial answers only)
NP	Number of trials in pre-test: the number of answers to the pre-test given by a student
SF	Final score: the final mark that was assigned by the examiner

The number of user actions Nu includes any selection or decision made during game play. It covers the retrieval of resources, opening of a video, entering a location and the navigation to other services. It thus includes some of the other variables. Locations are spaces in the game environment, for instance "rooms" where the players may find specific information. Resources include relevant papers, reports, letters or other documents, URLs, graphs etcetera that are made available in the game. Since the retrieved resources are mostly loaded (automatically) in a new browser window different from the game session window, only the opening of a resource is logged, while its closure is not. This means that the logging doesn't allow for extracting the time spent to resources. Videos can be any recorded file including expert interviews, instructional videos, archived TV-programs or documentaries. The logging data indicate exactly when a video was called and launched, but we were unable to detect pauses inserted by the user. We defined the time TV spent to a video clip as the duration from its launch to the next detected user action. This means that TV denotes an upper limit of the time spent to the videos, which may either be smaller than the actual duration of the video clip (if interrupted), or larger than its actual duration. Whenever a video clip is loaded, the value of TVL (which is the total duration of the videos that are accessed) is augmented with the duration of the clip. Hence TVL provides a reference that indicates the time required to view all the clips that were accessed. Pre-test scores SP are based on the initial answers of the students to the pre-test questions (40 items in total). Since students were allowed to change their answers, we used their initial answers to obtain a metric of prior knowledge. The final assessment scores SF are obtained from the examiners who assigned the scores on the basis of the reports about the respective games. It should be noted that the pre-tests and the final test are very different: the pre-tests are a simple check of basic knowledge required for entering the games, while the final test covers all contents covered by the games. This means the two aren't comparable and shouldn't be mistaken for a pre-test/post-test pair in an experiment. We used MS-Excel for filtering the log files of 118 students and used SPSS for further statistical processing. In the descriptive statistics of these key variables we have traced two outliers with z-score > 3. These outliers were kept in the sample.



6. Results

In this section we present the findings of the analysis that we carried out for addressing the respective research questions (RQs):

- RQ1 To what extent can we identify different gaming behaviours?
- RQ2 To what extent can behavioural characteristics be predictors of final scores?
- RQ3 Do we observe relationships between prior knowledge and gaming behaviours and performances?
- RQ4 How do students treat the video resources in the games?

6.1. RQ1 To what extent can we identify different gaming behaviours?

For addressing this research question we will present descriptive statistics of the key variables listed in table 1 and explore correlations between those.

6.1.1. Descriptive statistics of key variables

Table 2 summarises key figures of the logging analysis.

Table 2. Descriptive statistics of the logging data

	Average per student	Standard deviation	Coefficient of variation
Total time T	53.8 h	24.6 h	0.44
User actions N_u	5225	1891	0.36
Locations accessed N_L	156	61	0.39
Resources accessed N_R	180	96	0.54
Videos accessed N_V	120	42	0.35
Total time spent to videos T_V	7.4	5.7	0.77
Total length of videos accessed T_{VL}	5.6	2.1	0.37
Pre-test score S_P	6.6	1.2	0.18
Pre-test answers N_P	65	42	0.65
Final score S_F	6.6	1.5	0.24

The students' average of the total time T required for the 5 games is about 54 hours. The standard deviation of 24 hours indicates considerable spread among students. For example, the shortest time observed in the sample was 10.4 hours, the longest time observed was 146 hours; both students passed their exam (marks 6.5 out of 10, and 6.6 out of 10, respectively). Similar large variabilities are observed in the number of user actions N_u , the number of accessed locations N_L , the number of resources N_R , the number of videos N_V , and the number of pre-test answers N_P . It should be noted that the games provide access to a limited set of resources (89), locations (23) and pre-test questions (40). So, many of these assets are revisited. Location re-visits are often enforced, since some locations re-occur in all games and have to be re-opened in each game. The average number of videos accessed N_V is 120, which is considerably lower than the total number of 212 videos that are available in the game. This is because quite some of the videos are part of structured paths that use mutually exclusive video alternatives. Also, the video interviews include some less relevant issues, which players may want to skip. The average time spent to the videos T_V is 7.4 hours. This total length is substantially longer than the total net duration of the videos that were accessed (T_{VL} is 5.6 hours, not to be mistaken with the total duration of all videos available, which is 5.7 hours). The ratio T_V/T_{VL} is 1.32, which indicates that the players spent more time to the videos than needed to fully watch the videos: players may have paused the videos, may have made notes after completion, or - of course - may have been distracted by circumstances not related to the game.

6.1.2. Descriptive statistics of rate variables

Motivated by the variance of total time T , which indicates substantial differences in pace, we've also determined descriptive statistics for the time-based rates at which the various activities in the game were carried out. Table 3 shows the descriptive statistics of selected rate variables.

Table 3. Descriptive statistics of selected rate variables

	Average per student	Standard deviation	Coefficient of variation
Final score per unit time S_F/T	4.03E-05	2.22E-05	0.55
User actions per unit time N_u/T	3.10E-02	1.69E-02	0.54
Access rate locations N_L/T	9.08E-04	4.17E-04	0.46
Access rate resources N_R/T	9.93E-04	4.88E-04	0.49
Access rate videos N_V/T	6.86E-04	2.56E-04	0.37

As can be read from the coefficients of variation in table 2, the variability of behavioural rates are considerable. The variabilities tend to be equal or higher than those of the variables of table 1 (except for the access rate of resources, which is slightly lower).

6.1.3. Correlations between rate variables

In order to analyse the coherence of variability across different behavioural indicators we have calculated Spearman correlations between the rate variables of the final scores, user actions, accessed locations, accessed resources and accessed videos. Table 4 shows the correlations R of the selected rate variables.

Table 4. Spearman correlation coefficients R for the respective rate variables

	Final score rate S_F/T	User action rate N_u/T	Access rate locations N_L/T	Access rate resources N_R/T	Access rate videos N_V/T
Final score per unit time S_F/T	1	0.702	0.649	0.468	0.656
User actions per unit time N_u/T	0.702	1	0.697	0.406	0.494
Access rate locations N_L/T	0.649	0.697	1	0.500	0.576
Access rate resources N_R/T	0.468	0.406	0.500	1	0.484
Access rate videos N_V/T	0.656	0.494	0.576	0.484	1

All correlations are significant at the 0.01 significance level (actually, in all cases $p < 0.001$). Their magnitudes are between 0.4 and 0.7. The mean value of all non-diagonal correlations is 0.56, which is substantial. Considerable smaller correlation coefficients and less significance are found when the absolute variables (cf. table 1) are used instead of the time-based rates (cf. table 2). It demonstrates that all variables point at the same direction: students who display a high rate in one variable are likely to also have high values for the other rate variables. Statistically, observing a significant correlation coefficient R between two variables means that the variability of one variable explains R^2 of the variability of the other variable. For example: the correlation of 0.697 between the action rate N_u/T and the locations access rate N_L/T , means that the variability of the action rate N_u/T explains 0.49 (R-squared) of the variability of the locations access rate N_L/T . These results signal some behavioural consistency between the rates of learning gains, user actions, accessed locations, accessed resources and accessed videos. Both the correlation and the variance identify the students' "switching behaviours" as a likely behavioural characteristic. Apparently, the rate of switching between different assets in the game is a consistent and inherent personal trait.

We conclude that the data show a large behavioural variability across students. Correlation analysis suggest a behavioural trade that reflects the rate of “switching” between different game objects, in this particular case the switching between written resources, video clips, and game locations, if not any user actions per unit time.

6.2. RQ2 To what extent can behavioural characteristics be predictors of final scores?

Based on the above we have carried out a multiple regression analysis (hierarchical forced entry) to search for any relationship between the “switching rates” and the learning gains as derived from the final assessment scores. In view of the time-based activity rates we have expressed the learning gains as scores per unit time S_F/T , which represents the efficiency of learning. Table 5 shows the model estimates of the regression analysis.

Table 5. Models of hierarchical regression explaining learning efficiency S_F/T

Models of learning efficiency S_F/T	User action rate N_u/T	Access rate locations N_L/T	Access rate resources N_R/T	Access rate videos N_V/T	R^2	F
Model 1	0.001	0.007	0.003	0.030	0.632	48.5
Model 2	0.001	0.008		0.031	0.628	64.3
Model 3	0.001	-	-	0.035	0.619	93.4
Model 4	0.001	-	-	-	0.492	112.5
Model 5	-	0.022	-	0.037	0.540	67.5

All models that we explored display have high explanatory powers, demonstrated by the high values of R^2 , ranging from 0.49 to 0.63. In models 1 and 2 overfitting and collinearity are an issue (not displayed in the table), with degenerated eigenvalues and variance inflation indices larger than 2 for the resources rate N_R/T and locations rate N_L/T . This is understandable since the predictors are highly interdependent. In particular, the user actions N_u/T reflect an all-inclusive rate that comprises the other predictors as well as remaining navigation actions. In model 3 it turns out that the video access rate N_V/T is less interfering with N_u/T , which makes model 3 a plausible model. Model 4 is the most simple model as it only includes the number of user actions as a predictor. Unfortunately, it loses some explanatory power by omitting the video rate as a predictor. Finally, model 5 is obtained as a solution that doesn't include the user actions N_u/T as a predictor. Altogether, model 3 yields the best result. It means that switching behaviour as based on video access rates N_V/T and overall activity rates N_u/T is a partial predictor (62%) of learning efficiency. However, a high learning efficiency, that is, the final score obtained for the delivered report, divided by the total time spent to the game, isn't the same as a high final score. One may wonder: do students who spend more time to the games learn more and therefore achieve higher marks, or do they in contrast need more time because they have difficulties with the learning? The data, however, fail to provide any evidence for resolving this issue: the correlation between scores S_F and total time spent T appear to be too weak ($R=0.182$, $p=0.049$) to be meaningful. A combined model of switching behaviours and total time spent T for predicting final scores S_F failed to produce meaningful outcomes. Apparently, other factors are predominant in final scores. Yet, we found some evidence that switching behaviour can be used as a predictor of total time spent. This spreads light on the question: do fast switchers study faster, or is it the opposite, if there is any effect at all? Table 6 summarises the coefficients of the regression models for predicting total time spent.

Table 6. Models of hierarchical regression explaining final total time spent T

Total time spent T	User action rate N_u/T	Access rate locations N_L/T	Access rate resources N_R/T	Access rate videos N_V/T	R2	F
Model 6	-1.25E+06	-4.00E+07	1.65E+07	1.40E+08	0.480	26.1
Model 7	-1.75E+06	-	1.04E+07	1.56E+08	0.464	32.9
Model 8	-2.68E+06	-	-	1.49E+08	0.452	49.2
Model 9	-2.80E+06	-	-	-	0.309	51.8
Model 10	-	6.47E+07	-	1.42E+08	0.445	46.0

Collinearity and overfitting (not presented in table 6) turned out to disqualify the first two models. These problems are overcome in model 8 while hardly losing explanatory power. Model 9 is less powerful, while model 10, which removed N_u/T as an overall predictor, has similar potential as model 8. We may conclude that the variability of switching behavior, either defined by the predictor pair N_u/T and N_v/T , or by the predictor pair N_l/T and N_v/T explains up to 45% of the variability of playing time T. Model 10 predicts that fast switchers spent more time to the games. This may seem a bit counterintuitive, since it would mean that fast switchers aren't fast and efficient learners. An alternative interpretation could be that fast switching produces shallowness and superficiality, if not mental fragmentation, which are counterproductive to learning. It means that the player's attention span for an asset is relatively low, whereby many assets have to be accessed and re-accessed. For model 8 the interpretation is more complex, because the minus sign that appears in model 8 (table 6) indicates that the two predictors act in opposite directions. Further analysis of the model characteristics has shown that video access rates N_v/T are dominant over the overall access rates N_u/T . This means that the model predicts a longer gaming time T at higher switching rates, unless in the extreme case that the z-scores of N_v/T are much lower than those of N_u/T and the latter start dominating the model. A tentative interpretation: low video access rates may be an indicator of conscientiousness, perseverance, patience and dedication, which in all cases may turn out to be preconditions for efficient learning. More evidence need to be collected for this proposition. Based on the joint results of model 8 and model 10, we may conclude that fast switchers are likely (45%) to spend more time to the games.

With respect to RQ2 we conclude that switching behaviour as based on video access rates N_v/T and overall activity rates N_u/T can be used as a predictor of learning efficiency. The data didn't give rise to an appropriate model for predicting final scores S_F . We found slight evidence that students who display increased switching behaviours require more time to complete the games.

6.3. RQ3 Do we observe relationships between prior knowledge and gaming behaviours and performances?

We have calculated Spearman correlations coefficients of the pre-test variables (N_P , S_P) with various behavioural and performance variables (N_u , N_u/T , T, S_F). In most cases, correlations are absent or weak. One might hypothesise that students who need many efforts to pass the pre-test (N_P is large) are weak students who need more time (T) for completing the games. No evidence for this was found in the data. A weak correlation of $R=0.201$ between user actions (N_u) and pre-test-score (S_P) was found to be significant at the 0.05 level. Despite its significance, the result is too weak to be meaningful. We found only one correlation that was significant at the 0.01 level, which is the correlation between the number of user actions (N_u) and the number of pre-test answers given (N_P): $R=0.487$ ($R^2=0.237$). It is tempting to conclude that switching behaviours (cf. model 4) are an inherent personal characteristic that is already traceable during the pre-test as well as during the rest of the games. Similar results were found in a regression analysis after replacing N_u (cf. model 4) with the dyade (N_u , N_v), which is according to model 3 a more accurate indicator of switching behaviour. However, in both cases the explanatory power of the models ($R^2=0.237$ and $R^2=0.241$, respectively) remains very limited.

We conclude that we cannot confirm any relationship between observables from the pre-test and other variables of the games.

6.4. RQ4 How do students treat the video resources in the games?

Video is widely recognized as a powerful educational medium because of the impelling nature of its multi-modality, its fixed temporal properties and its potential of representing realism [36]. However, video's intrinsic linearity and uni-directionality may easily get students to passivity and loss of attention [37]. Also, creating purposeful videos is laborious, if not costly. Hence it is interesting to know how students deal with the 220 videos in de VIBOA games. Worst case, students might skip all the videos without watching them. In table 1 we explained that we were able to extract three video-related variables, which are the total number of video clips accessed by a student (N_v), the total time spent to watching videos (T_v) and the total length of the videos retrieved by the student (T_{vL}). Total time spent to videos was 7.4 hours (table 1). As compared with the total time spent to the game students spend 14.1% of their time to the videos. The standard deviation of 7.8% stresses that the students display a large variation of time spent to the videos. We recall that the ratio T_v/T_{vL} reflect to what extent the videos that were opened by the player are fully watched, or rather to what extent the time spent to the videos corresponds with the minimum time needed to watch the videos. The ratio of 1.32 indicates that the players spent more time to the videos than the total length of the videos that they retrieved. We should bear in mind, however, that the value of T_v is no more than an upper limit and that students may do other things during the time that videos are played. Still, what we do know from the logging is that students very rarely interrupt the videos untimely by performing some other user action. These observations provide evidence for the conclusion (which in accordance with the purpose and design of the games) that the videos are extensively used and consulted by students: they spend hours to them.

As a next step we explored correlations between diverse video-related variables (table 7).

Table 7. Significant Pearson correlation coefficients for various video-related variable pairs

	Number of videos accessed N_v	Average time per clip	Total time spent to videos	Fraction of total time spent to videos
Final score S_F	0.302			
Total time T	0.453	0.409	0.555	
User actions per unit time N_u/T	-0.253	-0.288	-0.332	
Final score per unit time S_F/T	-0.293	-0.251	-0.313	
Number of videos accessed N_v	1		0.570	0.295
Average time per clip		1	0.847	0.699

Although these correlations are all at the 0.01 significance level, the data aren't all that informative. The fact the average time per clip displays high proportionality with total time spent to videos ($R=0.847$) and with the fraction of total time spent to videos ($R=0.699$) is more or less straightforward, since these variables express to some extent the amount of effort put into the videos. Final score is weakly but positively correlated with the number of videos accessed ($R=0.302$). Likewise total time hints to be proportional to the efforts put into the videos. The negative signs in the correlations of S_F/T suggest that learning efficiency goes down when more effort is spent to the videos. This would be plausible because watching all these videos takes time. The same holds for user actions per unit time N_u/T (which indicates switching behaviour according to model 4): the more attention is paid to the videos, the less switching occurs.

In conclusion, the data about the videos suggest that students spend quite some time to the video materials. Overall, various plausible correlations between variables are observed, but despite their significance levels they are too small to draw definite conclusions and start building predictive models.

7. Discussion and Conclusions

In this study we've presented our analysis of existing student logs of the VIBOA environmental policy games at Utrecht University. These are inquiry-based serious games that offer students a lot of freedom of movement. Given this condition, we've observed substantial behavioural variability across different individuals. Such variability was revealed in diverse variables, including the number of accessed locations NL, the number of accessed videos NV, the number of accessed resources NR, the total time spent to the built-in videos (TV), the number of trials in the pre-test NP and the time T spent to the games. It turned out that the resources that are listed in the game's inventory (NR) were opened about twice. We note that this observation is a bit deceptive, however. All resources automatically open in a new browser window, which can then be re-consulted over and over again by the player without making the game engine aware of this. So, the ratio defines a lower limit. It is quite likely that resources are consulted much more often than twice, possibly in parallel with other tasks, e.g. watching a video, or navigating to another location. For the cross-correlations of various rate variables (SF/T, Nu/T, NL/T, NR/T, NV/T) we found consistent values between 0.4 and 0.7, all the 0.01 significance level. All variables point at the same direction: they suggest that the students' rates of switching between different assets in the game are a consistent and inherent personal trait. With respect to RQ1 (To what extent can we identify different gaming behaviours?) we conclude that the freedom of movement that the games provide, goes with a large variety of behaviours. As a main candidate behavioural dimension we've identified switching behaviour, which indicates the students' rates of switching between different assets in the game.

For addressing RQ2 (To what extent can behavioural characteristics be predictors of final scores?) we've used multiple regression analysis, constrained to linear models. We found that switching behaviour as based on video access rates NV/T and overall activity rates Nu/T can be used as a predictor of learning efficiency. Both the model and the model coefficients were found to be significant at the 0.01 level. The sample size of 118 subjects is well above the minimum requirements for applying regression analysis, according to Green's [38] rules of thumb (50+8k, and 104+k with k the number of predictors). This strengthens the reliability of the outcomes. The variability in these predictors explain 62% of the variability of learning efficiency. We also found some evidence that students with high switching rates require more time to complete the games. With respect to RQ3 (Do we observe relationships between prior knowledge and gaming behaviours and performances?) the answer is negative. Variables extracted from the pre-test data were not, or only very weakly correlated with other behavioural or performance variables. One might say that the purpose of the introductory game and the associated pre-test is to bring the students to the right level for entering the subsequent games. This suggests that once the students have passed the pre-test, they all start the next games with sufficient prior knowledge. This would explain that the data of the pre-tests are independent from the data in the rest of the games.

Regarding RQ4 (How do students treat the video resources in the games?) the data suggest that students spent quite some time to the video materials. Although the videos were certainly designed as essential components of the games and present relevant ideas and clues about the topics, developers might fear that students tend to skip these, because the linear, temporal and unidirectional nature of video puts heavy demands on students who might impatiently want to finish their work. Such fear appeared to be groundless: students spend hours watching the built-in videos.

Some final remarks aim to put our findings in perspective. In this study we had access to the existing log data and final scores only. We had to do without background profiles of students and couldn't make use of matched pre-tests and post-tests, questionnaires, direct observations, and a randomised trial with experimental groups and a control group. Therefore, our aim was not to test hypotheses or to construct predictive models, but to explore the phenomena and try to identify some of its determinants. So far, there was little theory that would support and explain the phenomena studied, which makes interpretation somewhat speculative and provisional. We have used key concepts such as "freedom of movement" and "switching behaviour" in an ad hoc way, but a next step would require proper definitions and better alignment with existing theories and concepts. Also, our study was inherently tied to the specific game contents, game designs, educational context and user groups of the VIBOA games, which doesn't necessarily produces general validity. Different contexts, different games or even different resources, videos, or locations in the very same games may produce different effects and relationships. Despite the limitations of our study, we have demonstrated the rich potential of analysing player log files and revealed some relevant phenomena and variables. Next steps in research would be to devise well-controlled experiments or quasi-experiments, while at the same time confronting learning analytics



with other techniques for observing playing behaviours. Since serious gaming are among the richest and dynamic online learning environments, they are destined to become an exemplary case of learning analytics.

References

- [1] VandeWalle, D., Brown, S.P., Cron, W.L., Slocum, L.W., “The influence of goal orientation and self-regulation tactics on sales performance: A longitudinal field test”. *Journal of Applied Psychology (APA)*, vol. 84, pp.249–259, 1999.
- [2] Fisher, S.L., Ford, J.K., “Differential effects of learner effort and goal orientation on two learning outcomes”, *Personnel Psychology (Wiley)* vol. 51, pp.397–420, 1998.
- [3] Redeker, C., Punie, Y., Ferrari, A., “eAssessment for 21st century learning and skills”, In Ravenscroft, A., Lindsteadt, S. Kloos, C.D., Hernandez-Leo, D. (eds.). *21st Century Learning for 21st Century Skills. Proceedings of the 7th European Conference on technology-enhanced learning EC-TEL*. Springer, Heidelberg, pp.292-305, 2012.
- [4] Westera, W., Nadolski, R., Hummel, H., Wopereis, I., “Serious Games for Higher Education: a Framework for Reducing Design Complexity”, *Journal of Computer-Assisted Learning (Wiley)*, vol. 24, no. 5, pp. 420-432, 2008.
- [5] Westera, W., Nadolski, R., Hummel, H., “Learning analytics in serious gaming: uncovering the hidden treasury of game log files”, Paper presented at the GALA Serious Gaming Conference October 24-25 in Paris, France, 2013.
- [6] Buckingham Shum, S., Ferguson, R., “Social Learning Analytics”, *Educational Technology & Society (IFETS)*, vol. 15, no. 3, pp.3–26, 2012.
- [7] Baker, R.S.J.D., Yacef, K., “The state of educational data mining in 2009: A review and future visions”, *Journal of Educational Data Mining* vol. 1, no. 1, pp.3-17, 2009.
- [8] Romero, C., Ventura, S., “Educational data mining: A survey from 1995 to 2005”, *Expert Systems with Applications (Elsevier)*, vol. 33, no. 1, pp.135-146, 2007.
- [9] Bienkowski, M., Feng, M., Means, B., “Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief”. SRI International U.S. Department of Education, Office of Educational Technology, Washington, 2012. Retrieved from <http://www.ed.gov/edblogs/technology/files/2012/03/edm-la-brief.pdf>
- [10] Zouaq, A., Jocsimović, S., Gasević, D., “Ontology Learning to Analyze Research Trends in Learning Analytics Publications”. In Mathieu d'Aquin, M., Dietze, S., Drachler, H., Herder, E., Taibi, D. (eds), *Proceedings of the LAK Data Challenge, held at LAK 2013, the Third Conference on Learning Analytics and Knowledge, Leuven, Belgium, April 9, 2013*. Available at http://ceur-ws.org/Vol-974/lakdatachallenge2013_08.pdf
- [11] Martin, T., Sherin, B., “Learning Analytics and Computational Techniques for Detecting and Evaluating Patterns in Learning: An Introduction to the Special Issue”, *Journal of the Learning Sciences (Taylor & Francis)*, vol. 22, no. 4, pp.511-520, 2013. DOI: 10.1080/10508406.2013.840466
- [12] Arnold, K.E., “Signals: Applying academic analytics”, *Educause Quarterly*, vol. 33, no. 1, pp.10-15, 2010. Retrieved from <http://www.educause.edu/ero/article/signals-applying-academic-analytics>
- [13] Pistilli, M., Arnold, K., “Course signals at Purdue: Using learning analytics to increase student success”, *2nd International Conference on Learning Analytics and Knowledge, Vancouver, 2012*.
- [14] Long, P.D., Siemens, G., “Penetrating the Fog: Analytics in Learning and Education”, *Educause Review*, vol. 46. No. 5, pp.31-40, 2011. Retrieved from <http://www.educause.edu/ero/article/penetrating-fog-analytics-learning-and-education>
- [15] Cooper, A., Powell, S., MacNeill, S., “Survey of the State of Analytics in UK Higher and Further Institutions 2013”, *White Paper 2013:WP03, Bolton: CETIS*. Retrieved from <http://publications.cetis.ac.uk/2013/884>
- [16] Kraan, W., Sherlock, D., “Analytics Tools and Infrastructure. CETIS Analytics Series, vol. 1, no.11, 2013, Bolton: JISC/CETIS. Retrieved from <http://publications.cetis.ac.uk/2013/535>
- [17] Hung, J.-L., Hsu, Y.-C., Rice, K., “Integrating Data Mining in Program Evaluation of K-12 Online Education”, *Educational Technology & Society (IFETS)*, vol. 15, no. 3, pp.27–41, 2012.



- [18] Greller, W., Drachsler, H., “Translating Learning into Numbers: A Generic Framework for Learning Analytics”, *Educational Technology & Society (IFETS)*, vol. 15, no. 3, pp.42–57, 2012.
- [19] Siemens, G., “Learning analytics: A foundation for informed change in higher education”, Educause, slides retrieved from <http://www.slideshare.net/gsiemens/learning-analytics-educause>
- [20] Elias, T., “Learning analytics: Definitions, processes and potentials”, retrieved from <http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf>
- [21] Campbell, J.P., Oblinger, D.G., “Academic Analytics”, Educause White Paper 2007, retrieved from <http://www.educause.edu/ir/library/pdf/PUB6101.pdf>
- [22] Dyckhoff, A.L., Zielke, D., Bültmann, M., Chatti, M.A., Schroeder, U., “Design and Implementation of a Learning Analytics Toolkit for Teachers”. *Educational Technology & Society (IFETS)*, vol. 15, no. 3, pp.58–76, 2012.
- [23] Boyd, D., Crawford, K., “Six Provocations for Big Data, A Decade in Internet Time”, Symposium on the Dynamics of the Internet and Society, Oxford Internet Institute, Oxford, 2012. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431
- [24] Bluemink, J., Hämäläinen, R., Manninen, T., Järvelä, S., “Group-level analysis on multiplayer game collaboration: how do the individuals shape the group interaction?”, *Interactive Learning Environments (Taylor & Francis)*, vol. 18, no. 4, pp.365–383, 2010.
- [25] Fernández-Gallego, B., Lama, M., Vidal, J.C., Mucientes, M., “Learning Analytics Framework for Educational Virtual Worlds”. In *Proceedings of the 2013 International Conference on Virtual and Augmented Reality in Education, Procedia Computer Science (Elsevier)*, vol. 25, pp.443-447, 2013. DOI: 10.1016/j.procs.2013.11.056
- [26] Koper, R., Tattersall, C., “Learning Design: A handbook on modelling and delivering networked education and training”, Springer, Berlin, 2005.
- [27] Gobert, J.D., Sao Pedro, M., Raziuddin, J., Baker, R.S., “From Log Files to Assessment Metrics: Measuring Students' Science Inquiry Skills Using Educational Data Mining”, *Journal of the Learning Sciences (Taylor & Francis)*, vol. 22, no. 4, pp.521-563, 2013. DOI: 10.1080/10508406.2013.837391
- [28] Westera, W., Hommes, M.A., Houtmans, M., Kurvers, H.J., “Computer-Supported Training of Psychodiagnostic Skills”, *Interactive Learning Environments (Taylor & Francis)*, vol. 11, no. 3, pp.215-231, 2003.
- [29] Kickmeier-Rust, M.D., Albert, D., “Micro Adaptivity: Protecting Immersion in Didactically Adaptive Digital Educational Games”, *Journal of Computer Assisted Learning (Wiley)*, vol. 26, pp.95-105, 2010.
- [30] Reese, D.D., Seward, R.J., Tabachnick, B.G., Hitt, B., Harrison, A., McFarland, L., “Timed Report Measures Learning: Game-Based Embedded Assessment”, In Ifenthaler, D., Eseryel, D., Ge, X. (eds.), *Assessment in Game-Based Learning: Foundations, Innovations, and Perspectives*. Springer, New York, 2013.
- [31] 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”, *Procedia Computer Science (Elsevier)*, vol. 15, pp.203-209, 2012. Retrieved from <http://www.sciencedirect.com/science/article/pii/S1877050912008344>
- [32] Shute, V.J., Ventura, M., Bauer, M., Zapata-Rivera, D., “Melding the Power of Serious Games and Embedded Assessment to Monitor and Foster Learning: Flow and Grow”, In Ritterfeld, U., Cody, M., Vorderer, M. (eds.). *Serious Games: Mechanisms and Effects*, Routledge, New York, pp.295-321, 2009.
- [33] Mislevy, R.J., Steinberg, L.S., Almond, R.G., “On the structure of educational assessment”, *Measurement: Interdisciplinary Research and Perspective (Taylor & Francis)*, vol. 10, pp.3-62, 2003.
- [34] Pearl, J., “Probabilistic reasoning in intelligent systems: Networks of plausible inference”, Kaufmann, San Mateo CA, Unites States, 1988.
- [35] Sloomaker, A., Kurvers, H.J., Hummel, H.G.K., Koper, E.J.R., “Developing scenario-based serious games for complex cognitive skills acquisition: Design, development and evaluation of the EMERGO platform”, in preparation, 2014.
- [36] Westera, W., “A didactic framework for audiovisual design”, *Journal of Educational Media (Taylor & Francis)*, vol. 24, no. 2, pp.87-102, 1999.
- [37] Koumi, J., “Designing Video and Multimedia for Open and Flexible Learning”, Routledge, London, 2006.

- [38] Green, S.B., “How many subjects does it take to do a regression analysis?”, *Multivariate Behavioral Research* (Taylor& Francis), vol. 26, pp.499–510, 1991.

