Examining Students' Behavior in a Digital Simulation Game for Nurse Training

Daria Novoseltseva [1], Catherine Pons Lelardeux[2], and Nadine Jessel [3]

[1]*IRIT, University of Toulouse, University Paul Sabatier*
Daria.Novoseltseva@irit.fr
[2]*IRIT, University of Toulouse, INU Champollion, Serious Game Research Lab*
Catherine.Pons-Lelardeux@irit.fr
[3]*IRIT, University of Toulouse, University Toulouse Jean-Jaurès*
Nadine.Jessel@irit.fr

Abstract

Digital educational games have evolved in recent years due to the need to support education and training focused on non-technical skills. Data gathered through interaction with the graphical user interface are explored and exploited to analyze the players' experience. Many researchers have pointed the importance to analyse players' in-game behavior, which can help to enhance the learning process, identify learners' strategies, and improve the effectiveness of the serious game. This study is devoted to the analysis of students' behavior in a simulation game called CLONE, which targets work scheduling, situation awareness, and decision-making. The students performance and their behavioral strategies are examined based on sequences analysis of players' in-game actions. Moreover, outlier detection is proposed as an instrument for obtaining information that might help better understand students behavior. The findings of the study show that such indicators as time spent on planning schedule, time spent on inspecting additional information, and intensity of delegation activity are significantly higher for successful games than for lost ones. The sequences analysis and clustering have revealed students' prevailing in-game strategies, which mostly consist of inspection, reading medical records, delegation, and scheduling. Eventually, outlier detection has disclosed the game sessions with uncertain strategies and unstructured scheduling.

Keywords: Nursing Simulation, Serious Game, Nurse Training, Game Analytics, Learning Strategies, Outlier Detection

1 Introduction

Nowadays educational applications aim at both improving learning quality in general and enhancing understanding of the learning process. In order to follow these two highlights, new types of educational environments have been developed in recent years. One of these environments that allows improving students decision-making skills and performance is a game [1]. The use of game with purposes apart from entertainment context is called serious game (SG) [2, 3].

The number of serious games as an educational tool has recently increased in many areas such as education, health, wellbeing, and government [4]. Through serious games players learn from their own experiences [5, 6]. They are designed to educate people about a certain subject, amplify concepts, reinforce development, or help learners to gain learning skills or change an attitude. Thus, serious games are relevant instruments for education purposes, delivering a message, or knowledge acquisition through interactivity, motivation, and engagement. This determines the parallel interests in both: (i) engaging features of games as an entertainment medium (ii) use games for learning.

Domain experts emphasize the challenging task of evaluating a game's outcomes and its positive impact [7]. First of all, this task includes the issue of finding features to assess a game's effectiveness. In terms of learning, two main emerging domains have been identified: Learning Analytics and Gamification Analytics. The Learning Analytics [8, 9] is the analysis and representation of data about learners in order to improve students' learning. It also intends to support teachers to improve their teaching and better understand the students' strategies. Recently, the field of Gamification Analytics has been invested by research [10, 11]. It describes methods and tools that help designers to understand the user's strategies and improve the gamification process. Gamification Analytics aims at defining metrics to monitor user interaction with a gamified system. Data collected through metrics are used to elaborate comprehensive insights and better understand the relationship between a gamified system and user's behavior. As an illustration, Heilbrunn *et al.* in [11] has proposed a data-driven monitoring and adaptation process within gamification project to monitor the gamification process. Dichev *et al.* in [12] has tried to mix the learning analytic approach and the gamification approach to encourage students to reflect and monitor their learning activities through a course gaming platform. A recent review of literature has shown that the game, used as interventional tools, can influence behavior by incentivizing, reinforcing, educating, providing feedback loops, prompting, persuading, or providing meaning, fun, and community. However, not all game elements will appeal to all consumers equally, and different elements might work for different people and in different contexts [13].

In our work, we focus on the analysis of behavioral tendencies in a digital simulation game to estimate its effectiveness and to understand users' behavior and limitations of game simulation. The analyzed dataset has been collected from the game-based simulation for students from Nursing Schools, which is called Clinical Organizer Nurse Education (CLONE). The study proposes a data analysis model with the implementation of data mining tools. Applying statistical criteria to features that indicate the game's outcomes, we have found out factors that determine the game's success. Using n-grams models, which have been explained in detail in [14], we have parsed students' sequences of actions during the game playing. Hence, based on obtained n-grams, we have accomplished clustering and outlier detection for inspection the main behavioral strategies during the game playing. The outcomes of the study might bring new insights in identifying in-game behavior that can evoke reinforcement of the learning through simulation games.

The paper is organized as follows: in Section 2 we consider the related works in assessing serious game, introduce CLONE as educational environment, and formulate the research questions; in Section 3 we observe algorithms and methods implemented in this work and data collection process; Section 4 presents experimental results; Section 5 and Section 6 discuss and conclude the study.

2 Research Frameworks

2.1 Assessing Serious Games for Learning

A wide range of applications has been developed in past decades due to the existing many advantages of using games as an educational environment [15–17]. However, domain experts often face the challenging task of how games should be evaluated for education and training purposes [7]. The assessment of serious games' effectiveness has been mainly studied in the following three dimensions: usability, engagement, and learning [18]. Here, usability determines the ease of use and learnability of the serious game, engagement reflects the ability of the game to engage users, and learning shows what learners will know or be able to do as a result of playing the serious game. According to Castellar et al. [19], there are three issues which cause the difficulties in the establishment of general effectiveness estimation process: (i) different activities in control-groups; (ii) different measures that are used for assessing effectiveness; (iii) different statistical techniques are used for quantifying learning outcomes. Consistent with the first issue, the effectiveness can be estimated through the comparison of pre- and post-game measurements. Although the tracking of the measurements' dynamics is indicative, differences in outcomes could be due to knowledge differences between individuals or groups. The second issue is induced by a variety of features that are used to assess effectiveness. The third issue addresses various techniques which are implemented to estimate effectiveness. In this work, we focus on the analysis of students' in-game behavior in order to assess such aspects as learning and engagement.

The choice of features for evaluation of game's effectiveness is a challenging task for domain researchers. Some of them have estimated effectiveness through engagement, motivation, and usability [20–23], others have examined the knowledge acquisition and skills gained through learning [24–27]. The analyzed features usually depend on the application purpose and the type of collected data. However, the most prevalent indicator is time spent on an in-game activity. Time, which students spend on finishing a task or completing a mission can indicate engagement, difficulty, lack of knowledge, unclear formulation of the problem. Cheng *et al.* in [28] have evaluated learners' behavior through analysis of time spent on viewing relevant materials embedded to the SG. Cagiltay *et al.* have measured the time spent on reading and responding answers in exploring the effects of game design, on learning outcomes and motivation [29]. Tellioglu *et al.* have used session time as an indicator of productivity in Crowd-Sourced Serious Games [30]. Along with time features, other numerical indicators are frequently examined in serious game analytics. These indicators comprise of multiple parameters such as a number of clicks, frequency of tool use, and duration of interaction, which can be interpreted as a specific indicator of player's behavior. For instance, Kerr *et al.* [31] have identified the main features of learners' proficiency by logging actions via mouse clicks during gameplay. Denden *et al.* have suggested that the learners personalities can be identified based on their actions and choices in the games and examine such indicators as earned during the game score and places that the learner visited while exploring the game environment [32]. Rowe *et al.* in [33] have used the gameplay log data to measure students implicit computational thinking practices. In our previous work, we have investigated factors that can impact game's outcomes: time spent on the game session, number of actions made during the game session, and some characteristics of repeated sessions [34]. In this work, we expand the analysis of these factors by considering additional game sessions' features, related to specific in-game activity: time spent on planning personalized care plan for patients; time spent on inspection - analyzing supportive materials, such as medical records or doctor's prescriptions, and the number of actions related to delegating activity.

Modern serious games can contain complex solution paths (open-ended), and allow players to make various actions to achieve the goals. The assessment of the skill-gaining within such SG by traditional measurements is a difficult task, which sometimes can not properly reflect the real situation. Therefore, many studies in the field explore various ways to track players' in-game behavior to better understand learning progress [35–38]. The most indicative data for analyzing players' behavior are sequences of actions committed during the game. Kang *et al.* [39] have emphasized several general techniques for tracing players' behavior. The first technique has spatial-temporal nature: the game's designers investigate players' behavior according to the exact locations of students for a specific time frame. The second technique is based on sequential pattern mining analysis. In this case, the frequent subsequences are identified in a condition of predefined values for the occurrence of patterns in data. However, these techniques have limitations such as the inability to check the time of events and the presence of loads of low-level meaningful actions. Thus, it is important to seek the approach, which corresponds to the aims of the research. For instance, Loh *et al.* [40, 41] have explored methods to find differences between sequences of actions of experts and novice users. Some other studies have clustered players into performance groups based on in-game actions [42–45]. Liu *et al.* [46] have used a multi-case approach to examine how behavioral patterns vary according to students' learning characteristics.

All these studies aim to enhance the learning process and improve an educational system. Nevertheless, they are focused on considering frequent and common patterns in data while infrequent patterns can contain relevant hidden information. Hence, along with analysis of the game's outcomes and students' behavioral strategies we have detected learners with rare sequences of actions or learners with uncommon characteristics. We call these learners - outliers, and the process of finding them - outlier detection. The importance of outlier detection is caused by the fact that outliers in a dataset can bring significant information for application domains. Outliers exist in almost every dataset and might be summoned for a variety of reasons. They can appear because of human error, instrument error, natural deviations in populations, fraudulent behavior, changes in the behavior of systems, or faults in systems [47]. Many data mining algorithms try to minimize the influence of outliers or eliminate them altogether. However, this could lead to the loss of important hidden information since one persons noise could be another persons signal [48]. Therefore, outliers can be relevant, especially for cases that result to make a decision based on data analytics, such as an education. Despite the relevance of outlier detection in educational data, this topic is not too common. According to Aldowah *et al.* [49], only 2.25% of investigated papers are devoted to outlier detection. As far as we know, in the Serious Game field, there are no studies dedicated to outlier detection.

2.2 CLONE as Educational Environment

The importance of designing educational environments and educational programs which reproduce the professional environment, is pointed by many researchers [50–53]. Thus, Flin *et al.* highlighted the importance of non-technical skills in the work of an anaesthetist [1]. Non-technical skills are composed of two categories: interpersonal skills and cognitive skills. Interpersonal skills, e.g. interact with others, leadership, and coordination, are skills which enable a team to solve complex problems [54]. Cognitive skills are used to organize tasks, manage complex situations and make decisions. Therefore, developing inter-professional simulations is a perfect way to train medical specialists.

In healthcare field, a large majority of digital training are dedicated to medical specialists and focus on technical skills such as surgery or emergency first aid. Very few of them were designed to train inter-professional teams [55, 56]). This is the case of ”3D Virtual Operating Room” [55] which intends to train in real-time the surgeon, the anesthetist and the operating nurse. The simulation provides real-life like professional situations designed to show the importance of communication, leadership and decision making skills. While a large majority of digital training simulations target medical staff, they gradually appear for nurse training all over the world [57]. This is the case of a serious game designed by Petit *et al.* [58] which have intended to improve nurses' clinical reasoning and detection skills in home-care and community settings. Similarly, Kilmon *et al.* [59] have designed an immersive virtual environment for nurse training and focus on speed and accuracy of nurse response in emergency situations requiring cardiopulmonary resuscitation. Most of them target technical skills.

The game-based simulation called CLONE is also dedicated to nurses while it intends to develop non-technical skills. CLONE is a software which provides various real-life-like professional cases. Nurse students are supposed to develop skills such as work organization, decision making, and situation awareness. This digital environment includes game mechanisms and interactive features such as task scheduling or shifting and decision-making. The designing process of CLONE contained three steps: the domain analysis, the human activity modelling, and the scenario, which are described in [60]. Fig. 1 presents the graphical user interface.

Figure 1: *The graphical user interface*

The player chooses a study case from the library of educational cases, which shortly describes the actual and expected situations in a virtual medical unit. A study case provides interactions that allow the players to complete the mission. Each case contains patients, actors involved in the medical team and medical dynamic events predictable or not. In each case, the player takes the role of a nurse and must program patients care for herself/himself and their teammates and make decisions under conditions of uncertainty and risk. The mission includes locks (educational and playful), which intend to prevent the player to succeed. The player has to manage patients' diseases to deliver care and to delegate tasks to the nurse assistant and nurse taking into account the workload of each actor of the game as well as their possibilities. To complete the mission, the medical care team (nurse and nurse assistant) has to provide the required care for each patient according to their pathology. Afterwards, outcomes are compared to expected objectives, and results of the game are immediately displayed on the dashboard.

The playing process includes the following steps:

1. **The briefing.** At this step a student reads information about the mission of the game. The game briefly describes the actual situation and expected situation at the end.

2. **The communication with the night shift.** Here, a student receives information from the night shift about the current situation when they shift at 6:30.

3. **The scheduling.** This step is devoted to developing a care plan for all patients. At this step, a student inspects patients' records and organizes their daily activities.

4. **The care delivery.** The main goal of this step is to provide drugs, organize medical examination, professional phone calls, patient discharge or arrival.

5. **The communication with the afternoon shift and debriefing.** The next shift is informed about the current situation.

A list of soft and hard constraints is attached to each patient. A hard constraint must be satisfied at all costs. For example, distributing an anti-inflammatory treatment according to a doctor's prescription is a hard constraint. A soft constraint refers to a desirable practice that might be violated in order to generate a workable solution. If a player breaks a soft constraint, the patient's health is not strongly affected. If a hard constraint is broken, a player loses a star. A set of stars is associated with each scenario and represents the maximum of allowed violations. Finally, when a student launches a new session and restarts a scenario, the game saves the global care plan from previous sessions and a student continues playing. Thus, in case of the failure, a student can try to play again using information and experience gained during the previous game sessions.

2.3 Research Questions

To examine the students' behavior while they interact with the game environment, this study addresses the following research questions:

RQ1: Which factors impact the game's outcomes?

Answering this question, we have considered game sessions from 2 different angles: comparison of features for successful and lost sessions, and comparison of results of repeated sessions when students play several times in a row.

RQ2: Which strategies did students choose to achieve the games purpose?

Here, we have referred to tendencies in students' behavior by highlighting frequent patterns of actions during playing.

RQ3: Can the detection of outlying sessions help domain experts to understand students' behavior better?

In the framework of this question, we have indicated the main reasons for the arising of outliers and the features that determine them.

In responding to these research questions, we have aimed to assess the game in terms of two aspects: learning and engagement. The prevailing behavioral tendencies might show the depth of involvement and learnability, while outlier detection might reflect limitations in learning. Consequently, this can help domain experts to understand learners and improve the learning process by provision data-based decision-making.

3 Methodology

One of the challenging tasks in-game analysis is to determine appropriate techniques, which correspond to the purposes of research. The choice of methods mainly depends on data type and data context. To answer formulated research questions, we have accomplished complex data analysis with the implementation of various data analytics and data mining techniques, such as non-parametric statistical analysis, sequences analysis, clustering, and outlier detection.

3.1 Methods

Statistical Analysis and Visualization

In SG analytics statistical analysis is closely related to data visualization. The visualization as a method of analysis has emerged in recent years as a new tool for tracking students learning progress and presenting students knowledge understanding [16]. Visual representation of data can reinforce recognition of playful learning patterns, as well as improve game design for optimal learning and engagement. The combination of statistical and visual methods can take the form of descriptive statistical charts, simple learning curves, histograms, and heat maps.

To address the first research question, along with basic descriptive statistics, we have utilized non-parametric statistical analysis for independent variables and descriptive statistics. In the context of visualization, we have represented outcomes as boxplots.

Sequences Analysis

The data extracted from the game represent sequences of actions, which students have made during the game session. For a symbolic sequence, the simplest way is to treat each element as a feature. However, the sequential nature of sequences cannot be captured by this transformation. To keep the order of the elements in a sequence we have exploited the n-gram models. N-gram approach is commonly used in computational linguistics (e.g. natural language processing) and computational biology (e.g. DNA sequencing or protein sequencing). An n-gram is a contiguous subsequence of n elements from considered sequence. For example, for a sequence of elements {1, 2, 3, 4} we can form three 2-grams {12, 23, 34} and two 3-grams {123, 234}. Given a set of n-grams, a sequence can be performed as a vector of the presence and the absence of the n-grams or as a vector of the frequencies. These vectors form map matrix, which afterward might be a basis for further analysis such as clustering or classification [14]. At the same time, n-grams can be used for efficient approximate matching [61].

In this work, the sets of n-grams have been utilized for computing a similarity measure between sequences of actions. We have applied a metric based on Jaccard index, which reflects ratio Intersection over Union.

$$J(A, B) = A \cap B / A \cup B \tag{1}$$

Here the numerator is the number of n-grams intersection and the denominator is the number of n-grams union. The Jaccard index ranges from 0 to 1, where 1 means 100% similarity and 0 means 0% similarity. Therefore, distance based on Jaccard index can be represented as follows:

$$dist(s1, s2) = 1 - J(s1, s2) \tag{2}$$

Where s1, s2 are sets of n-grams for sequences of actions.

Clustering

Among the data science techniques for exploration of behavioral strategies or types of game play, the most widely used are cluster techniques. Cluster analysis helps researchers to group data based on the similarity of the data points.

There are numerous options for clustering methods, but all of them attempt to merge similar objects in the same cluster, and split dissimilar objects into different clusters. For instance, Kerr and Chung in [31] have applied fuzzy feature cluster analysis to identify key features of student performance in log data collected from a mathematical game for sixth grade students; Slimani *et al.* in [43] used Expectation-Maximization and K-means clustering approaches for students' performance in a game-simulation of the biological room; Martin *et al.* in [62] have performed hierarchical clustering in order to investigate student success study of fractions.

In this work, we have implemented clustering method with a distance matrix based on Jaccard index in order to find out groups of students with similar sequences of actions and reveal their characteristics. Our main hypothesis has been founded on the assumption that clusters with small number of members contain outliers and their characteristics are abnormal. The partitioning methods of clustering have not been used due to the absence of predefined parameters except distance matrix. Eventually, we have applied agglomerative clustering as a linkage method, namely Wards method, which is based on minimizing variance. Obtained clusters have been considered and their characteristics have been compared.

Outlier Detection

In the context of analyzed data, we have assumed that the games where students had abnormal and rare sequences of action are called outliers and the process of finding them is outlier detection. Since the similarity measure between sequences has been based on a distance matrix (Jaccard distance), we have considered distance-based outlier detection technics.

One of the most commonly used distance-based approaches is k-nearest neighbor outlier detection (kNN). The kNN outlier detection algorithm finds outliers in data relative to their neighborhood. The basic idea is that outliers are data points that are distant from their neighbors or which have sparse neighborhoods. The algorithm is described by [63] and its steps are as follows: first, a k-nearest neighbor graph is defined, where every vertex has k edges to the k nearest vectors. The weight of the edge is the distance between vectors. Hence, according to a predefined threshold for edge weight (distance between points), instances are detected as outliers if at least one point in their k-neighborhood is further than the threshold.

kNN algorithm works well for detecting global outliers - objects which are outlying for the entirety of the dataset in which it is found. The results of the algorithm are very simple to understand and equally easy to interpret. The algorithm requires predefined parameters: k - number of nearest neighbors, d - the threshold. In the current work, we have considered k = 3. As a thumb rule, in the outlier detection field, the partition of the outliers usually does not exceed 10%; however, it depends on the application domain and other circumstances related to the certain context. Therefore, the assumption that 10% of students are outliers have been implemented, i.e. 10% of sequences with the highest value of outlying score have been defined as outliers.

3.2 Data Collection

In this work we have analyzed log data that include information about actions of students during game sessions and their timestamps. It is worth mentioning that we have not focused on time series and have not considered the time for each action. However, we have concentrated on continuous sequences of actions and their row in the game session. In other words, we have not compared data objects according to time, when they happened, but have analyzed their order.

The actions or sets of actions have been denoted as activity types and after have been coded. The code for each action had the following representation: Type_Info_Time, where Type corresponds to activity type, Info refers to the Patient ID or extended information about activity, and Time corresponds to the virtual time at schedule panel. The scheduled panel contains six time gaps with one-hour duration and one time gap with thirty minutes duration, where the start time is 7:00, and the end time is 13:30, thereby the code for gap 7:00-8:00 is 1, and code for gap 13:00-13:30 is 7. Depending on the activity, the code may consist of one, two, and three parts. For example, the code of such activity as Communication is C, the code of planning of the breakfast Pm_bkf, the code of planning an activity pertaining to personal care for a Patient 4 at 8:00-9:00 time gap is P_4_2. The full list of codes and its descriptions are represented in Table 1.

Table 1: *Types of actions*

Type	**Description**	**Code**
Care action	Deliver care to a patient	A_1/.../A_5
Related activity	Perform tasks related to patient's care	Af_1/.../Af_5
Relaxation	Perform actions in the relaxation room	R
Delegate tasks	To a nurse assistant To a nurse To a nurse and nurse assistant	D_AS D_IDE D_IDE-AS
Planning Micro	Set / modify / delete a task in the patient's care plan	P_1_1/ .../P_5_7
Planning Macro	Move/ delete a task in the global care plan	Pm_1_1/ .../Pm_5_7
Arguments	Argue for the choice	Ar_1/.../Ar_5
Open patient's records	Open patients administrative record and click on mask to discover hidden information Open patients medical record and click on mask to discover hidden information Open patients nursing record and click on mask to discover hidden information	O_adm O_med O_nurse
Communication	Communication with the next or previous shift team	C
Help	Get information about the nursing practices in this medical unit	H
Inspection of global planning	Inspect scheduled global planning without making any actions	PI
Inspection	Inspect a patient's care planning without making any actions for patient Inspect administrative record Inspect nurse assistant record Inspect all information Inspect global care plan for all patients Inspect medical reports Inspect nurse service and nurse assistant service Inspect nurse service Inspect prescriptions for injections Inspect prescriptions per os Inspect other requirements Inspect other information	I_1/.../I_5 I_adm I_assist I_info I_all I_med-rep I_n-as I_nurse I_pr-inj I_pr-os I_pr I_service

Thus, the collected data contained in-game actions for 353 game sessions which have been made by 222 students. The students from 11 Nursing Schools played the same scenario during the 2018-2020 period. The explored scenario included 5 patients' profiles, who require a low-level of care. According to the scenario, students had to attentively analyse patients profiles, doctor prescriptions, and schedule care plan for all patients (except Patient 3, who arrives later in the morning). Moreover, the set of established constraints should be satisfied for each patient. In case of breaking of hard constraints, students accumulate critical errors, which can bring to the failure of the game session. The maximal allowed number of critical errors for

Patient 1 and Patient 2 is 7, for Patient 3 and Patient 4 is 3, and for Patient 5 is 4. Among considered in this study 353 game sessions, 261 are lost, and 92 are successful.

4 Results

RQ1: Which factors impact the game's outcomes?

RQ1 has been examined by 2 steps analysis: (i) analysis of dependencies between the game's outcomes and game's features; (ii) analysis of repeated sessions and their features. In our previous study, we have found out that students, who spend more time on playing and make more actions per session achieve better results. In the current work, we have extended the list of factors that might impact game outcomes and have considered such indicators as time spent on planning schedule, time spent on inspection, and a number of actions related to delegating activity. Since analyzed indicators significantly differ from the normal distribution ($p<0,0005$), 3 statistical hypotheses have been checked by the non-parametric Mann-Whitney U test to find out the difference between successful and lost sessions.

H1. The game outcomes depend on the time spent to Planning schedule

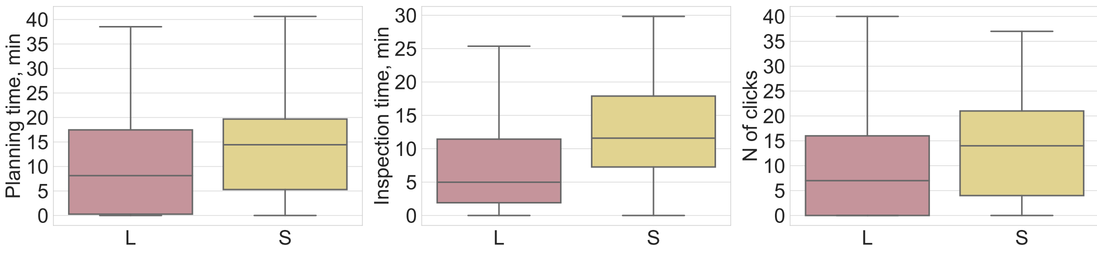
The purpose of the game is to schedule a day for patients and to deliver care. Therefore, the time, which students spend on the planning schedule reflects their engagement in the game process. The non-parametric Mann-Whitney U test for independent variables has shown the significant difference ($p<0,0005$) between the planning time of the successful session and the planning time of lost games (Fig. 2a). The average planning time of successful game sessions is 14 minutes (median = 14 minutes), whereas the average planning time for lost game sessions is 10 minutes (median = 8 minutes). Successful sessions require more time for planning than lost sessions. Therefore, we can accept the hypothesis that success depends on the time spent on the planning schedule.

H2. The game outcomes depend on time spent on Inspection

To follow purposes of the game, students need to gather and analyze additional materials, such as medical records, doctor's prescriptions, etc. All these activities refer to the same activity type, called inspection, which reflects students' preparation to decision-making (Table 1). Therefore, time spent on inspection corresponds to the time spent on exploring of information distributed through the virtual environment. Mann-Whitney U test has shown that the time spent on inspection for lost sessions is significantly lower ($p<0,0005$) than the time spent on inspection for successful games (Fig. 2b). The average time spent on inspection for successful games is 13 minutes (median = 12 minutes), whereas the average time for lost games is 8 minutes (median = 5 minutes). Here, we can outline that the more students inspect the more successful outcomes they have.

H3. The game outcomes depend on delegating activity

The last hypothesis is dedicated to delegation activity which determines the ability to assign a task to other teammates. There are three options in this activity type: assign a task to another nurse, assign a task to a nurse assistant, and assign a task to both nurse and assistant to accomplish it in pairs (Table 1). Student has to take into account the workload of each actor of the game as well as their possibilities. For instance, nurse assistance can not do some tasks, which nurse can. Therefore, delegation activity is balancing task, where student has to follow proposes according to game conditions. To estimate the delegating activity, we have collected a number of actions related to the delegation. Consistent with the Mann-Whitney U test delegation activity for lost sessions is significantly lower than for successful sessions (Fig. 2c). The average number of actions related to delegating for successful games is 14 (median = 14), while for lost games this indicator is 9 (median = 7). The proposed hypothesis can be accepted.



a) b) c)

Figure 2: *Boxplots of investigated features versus game's outcomes, where L corresponds to lost sessions and S corresponds to successful sessions. A box denotes the distribution of values between 25th and 75th percentiles. The line at the center is the median. a) Planning time. b) Inspection time. c) Number of delegation clicks*

According to Fig.2, there are game sessions where students have not not committed delegating activity either inspection or planning activities. It is totally normal because the majority of such sessions are repeated. When a student launches a new session and restarts a scenario, the game saves the global care plan from previous sessions and a student continues playing. Additionally, during repeated sessions students can gain information from previous game sessions. Consequently, the impact of these factors on the game's outcomes have been examined by analyzing repeated sessions. Among 222 students, who played the game, 81 students played more than one session and 21 students played more than two sessions (Table 2). We have already ascertained in [34] that with increasing order of session, such features as average session time and number of actions made per session are decreasing. This can be caused by the fact that game saves modifications of the schedule panel from previous sessions. Therefore, playing the second and the third time students spent less time on planning the schedule. Thus, planning, inspection and delegating activities have been also examined. Analogically with previous results, the average values of these features are decreasing when students play more. Despite this fact, the percentage of lost games stays relatively high: for the first sessions is 94%, whereas for the second sessions this value is 75% and for the third session is 76%.

Table 2: *Characteristics of repeated sessions*

Characteristic	1^{st} **session**	2^{nd} **session**	3^{rd} **session**
Number of players	81	81	29
Percentage of lost sessions	94%	75%	76%
Average time of session, min	50	23	14
Average time spent on Planning schedule, min	17	2	1
Average time spent on Inspection, min	10	7	2
Average number of actions per session	232	59	35
Average number of Delegation actions per session	14	2	1

Since the game saves modifications of the schedule panel from previous sessions, repeated sessions can be considered as continuous games. For example, a student may devote the first session to the inspection of patients' records and scheduling, and the second session - to deliver a care, which makes a comparison of sessions incorrect. To avoid biased results, repeated sessions for one student have been merged into one. Hereby, for further analysis, we have obtained 222 sequences of actions for 222 players.

RQ2: Which strategies did students choose to achieve the games purpose?

To examine the students' activity during the game, 222 sequences of actions have been parsed into n-grams. We have examined three models where parameter n = [4,5,6]. The models, where n $<$ 4 have not been considered because they can not show the main tendencies in game strategies of students. The game does not have any restrictions in terms of continuous progress, which means students can commit any action at any time and repeat it several times. As an example, students can plan an activity related to care for Patient 1 for 7:00-8:00 time gap, then for 10:00-11:00 time gap, and afterward modify 7:00-8:00 time gap. Two-grams and three-grams models contain a lot of noisy patterns, which have high frequency but do not reflect any behavioral features. In selecting an appropriate model for further analysis, we have focused on the comparison of the following model's characteristics: number of unique grams and relative frequency. The Relative Frequency RF for the gram G_i is a ratio of the grams frequency $N(G_i)$ (how many times the gram appears in data) and the number of existing grams in the data N:

$$RF(G_i) = N(G_i) \times 100\%/N \tag{3}$$

With the increasing n parameter of the n-gram model, the number of unique grams rapidly growing due to the expanding the number of possible combinations of actions (Table 3). Consequently, the part of grams which occur in data once is increasing. Therefore, for further analysis, the four-grams model has been chosen, due to the optimal values of considered characteristics.

Table 3: *Characteristics of n-gram models*

The model	4-grams	5-grams	6-grams
N of unique grams	22 041	32 126	40 395
RF of 100 most frequent grams	21%	14%	9%
RF for grams occurred in data once	26%	42%	56%

Patterns consisting of four actions can determine tendencies in players behavior during the game session. For example, the four-gram {P_1_2,P_1_3,P_1_4,P_1_5} shows continuous planning strategy and the four-gram {I_1,P_1_1,I_1,P_1_2} refers planning with inspection. Consistent with analysis of obtained frequent grams, the evident behavioral strategies in game sessions procedure have been highlighted (Table 4). In particular, the distinguished strategies concern the following activities: scheduling personalized care plan (S_P), delivering care (S_A), opening medical records (S_O), and inspection (S_I). Moreover, general strategies pertaining to inspection, opening medical records, and scheduling, have been specified consistent with patient index. Patterns devoted to delivering care have been merged into union strategy S_A, due to the small values of RFs for patterns related to a certain patient. According to analysis of patterns' frequency, the majority of patterns is devoted to the planning of schedule for patients. The scenario conditions imply focusing more on one group of patients and less on others. This strongly depends on the number of hard constraints for each patient and their discharge or arrival (according to the scenario, Patient 3 has late arrival). The distribution of patterns among strategies of planning, as well as strategies of inspection and reading medical reports have the same characteristics: the highest RFs correspond to Patient 1 and Patient 2 and the lowest to Patient 3 (Fig.3).

Table 4: *Strategies*

Strategy	**Description**	**Examples**
S_P_1, S_P_2, S_P_3, S_P_4, S_P_5	Planning a schedule for the certain patient	{P_1_2,P_1_3,P_1_4,P_1_5} {P_4_3,P_4_4,P_4_5,P_4_6} {P_5_1,P_5_2,P_5_1 P_5_2}
S_A	Delivering cares for patients	{A_1,A_2,A_4,A_5}
S_O_1, S_O_2, S_O_3, S_O_4, S_O_5	Opening medical reports and planning a schedule for the certain patient	{O_adm,O_med, P_1_2,P_1_3} {O_nurse,P_5_2,O_med,P_5_3}
S_O	Continuous opening medical reports	{O_adm,O_med,O_nurse,I_info}
S_I_1, S_I_2, S_I_3, S_I_4, S_I_5	Inspection and planning a schedule for the certain patient	{I_all,I_1,P_1_1,I_all} {I_2,P_2_1,P_2_2,P_2_3} {I_all,I_5,I_info,P_5_1} {I_2,P_2_1,I_2,P_2_2}
S_I	Inspection of additional information	{I_pr-inj,I_pr,I_med-rep,I_n-as} {I_pr-os,I_pr-inj,I_pr,I_n-as}

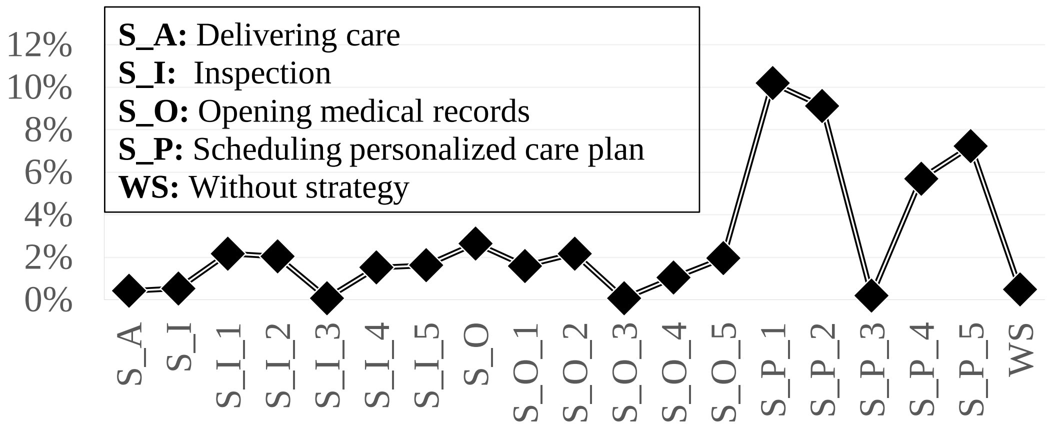


Figure 3: *Cumulative Relative Frequency of four-grams according to distinguished strategies. The figure concerns the most frequent 1281 four-grams with RF>0.01%. Here strategies related to inspection, opening medical records, and scheduling are specified for each patient and represented like **S_Action_n**, where n is the index of the patient. WS denotes patterns without strategy*

In order to emphasize the similarities between sequences of actions, which students made during a game session, a clustering have been exploited. Agglomerative hierarchical clustering (Wards method) has been implemented by using computed earlier Jaccard distance matrix for sequences of four-grams. The hierarchical dendrogram has divided sessions into four major clusters. The main characteristics of clusters are presented in Table 5, where N denotes the number of game sessions in the cluster, Min and Max show the minimal and maximal length of four-gram sequence in a cluster, Average determines the average number of four-grams per session in a cluster. The biggest cluster C4 contains 111 game sessions, whereas the smallest includes 18 game sessions.

Table 5: *General characteristics for clusters*

Cluster	N	Min	Average	Max
C1	67	109	302	467
C2	18	1	61	313
C3	26	77	227	386
C4	111	52	307	708

To investigate prevailing strategies for clusters, we have considered the distribution of patterns occurring in every cluster depending on their strategies (Fig. 4a). Distributions of clusters C1, C3, and C4 have similar shapes: patterns mostly distributed between planning, inspection, and reading medical records strategies. Meanwhile, the major part of patterns from cluster C2 has either an inspection strategy either does not have any strategy. More detailed distributions of patterns according to their strategies and clusters are represented in the Fig.4b (patterns without strategy are excluded). Here we can see the dependence between the number of hard constraints and the frequency: the most frequent patterns correspond to Patient 1 and Patient 2, whereas the less frequent patterns correspond to Patient 3.

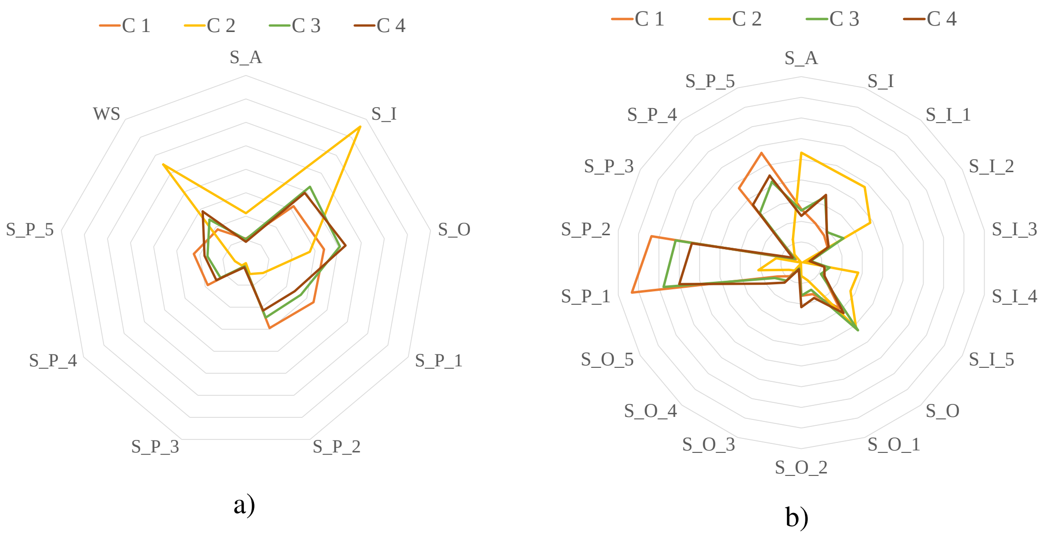


Figure 4: *Distribution of 4-grams by clusters. a) Distribution of patterns according to their strategies. Here WS denotes RF of patterns without strategies. b) Detailed distribution of patterns according to their strategies. Patterns without strategy are excluded.*

RQ3: Can the detection of outlying sessions help domain experts to understand students' behavior better?

Cluster C2 notably contrasts with other clusters in terms of patterns distribution by strategies. This means there are sessions, where players' behavior differs from the general behavior of the majority of players. To investigate the nature and significance of these differences, unsupervised outlier detection has been used . Twenty-two game sessions have been detected as outliers by applying the kNN method. Among them, four sessions are from cluster C4 and eighteen sessions are from cluster C2. Thus, the entire cluster C2 is outlined.

The distribution of action types for both classes of sessions - Inliers (Class N) and Outliers (Class O) is depicted in Fig. 5. Game sessions for Class N have the homogeneous spreading of actions, which determine the main behavioral tendencies for the majority of players. These game sessions are mostly devoted to planning schedules, reading medical records, and information inspection. During these game sessions, players use the delegation of tasks and deliver care. Contrary to this, sessions-outliers do not have any evident tendencies. However, we can emphasize several reasons for their arising:

1. **Delivering cares before planning.** Among detected outliers, there are games where students do not plan a personalized schedule for each patient before delivering care. As a result of this activity, a student makes critical errors that lead to failure.

2. **Inconsistent planning.** Elaborating a global care plan is one of the most important goals of the game. The majority of students try to consistently plan the daily care for one patient: inspecting information, reading medical records, setting or modifying personalized care, and afterward, they do the same actions for another patient. Furthermore, they effort to continuously move from the first time gap to the last. This group of detected outliers is characterized by inconsistent planning when students planned schedule just for one patient or they incoherently moved from one time gap to another.

Inconsequent planning interferes students to achieve the goal of the scenario, however, sometimes this tactic leads to success. Among detected outliers, there is one successful session, where the student often interrupted the inconsistent planning by inspection or delegation, that did not break hard constraints.

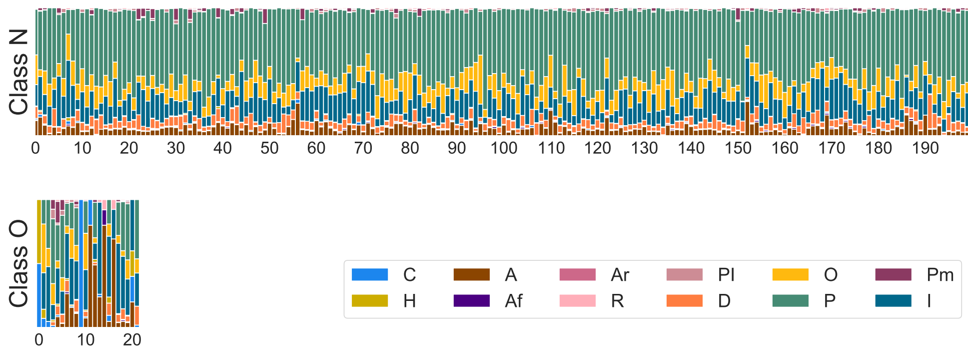


Figure 5: *Distribution of types of actions according to the class, where Class N contains inlier session, and Class O contains outlier sessions.*

5 Discussion

Serious games dedicated to learning have evolved in recent years as tools for gathering professional expertise [50–52]. The analysis of in-game students' behavior enables to evaluate skills-gaining through playing and assess game's effectiveness. In this work, we have investigated students' behavior in a game-based simulation CLONE, which intends to develop skills such as task management, situation awareness, and decision-making according to Flin's taxonomy [1].

Students from 11 French Nursing Schools played the same scenario, where they had to deal with 5 patients, who required low-level care. Inspecting medical reports and checking patients' profiles, students had to plan a nursing day and to deliver care for all patients. The work leaded by Novoseltseva et al. [34] have been extended by the analysis of additional game session features. Like others research papers [28–30], the present work considers time features, which indicate characteristics of playing. From the analysis of which factors impact the game's outcomes (RQ1), we have found out that successful games required more time for planning personalized care plans for patients. This indicator shows the involvement of students in the game process through interaction with the game environment. Therefore, we can infer that the more students are engaged, the more time they spend to gather information and memorize it. As a consequence, they achieve better outcomes. Along with scheduling and delivering care, students may inspect additional information, such as medical records or patients' profiles. This activity may reflect the preparation of students for decision-making and their situation awareness. According to the tested hypothesis, the time spent on inspection for successful sessions is significantly higher than for lost sessions. This proves that gaining information before decision-making improves the game's result: the more students are aware about the patient's pathology and required care, the easier it is for them to schedule a fitted personalized care plan. Finally, the delegating activity has been studied as a factor that impacts the game's success. During the game, students can assign a task to the nurse assistant, to the nurse, or to both nurse and assistant to do it in pairs. This is the balancing task which depends on actors' workload and possibilities (e.g. nurse assistance can not do some tasks, which nurse can). Drawing on the results of statistical test, we can conclude that successful sessions are characterized by more intense delegating activity than lost sessions.

To examine learning progress during the game, repeated sessions have been analyzed. We can emphasize that the more students play, the less time they spent to schedule, inspection, delegation. However, generally, students slowly progress during the second and third sessions. 75% of students who played the second time did not achieve the mission, whereas for students who played the third time this value is 76%. This means that students did not manage to find a good strategy or faced to the same difficulties without finding any solution.

Examining frequent action patterns has revealed the main in-game behavioral strategies, which students stand by (RQ2): inspection, opening medical records, scheduling personalized care plan, and delivering care. The most frequent four-grams indicated dependence between the cumulative relative frequency for the activity type and the level of care required by the patient. The analysis of relative frequency show that the patterns related to schedule care, opening, and inspection records for Patient 1 and Patient 2 are more frequent than for Patient 3. The clustering analysis reinforced the analysis of frequent patterns: members of three biggest clusters (C1, C3, C4) adhere to the same strategies - scheduling care plan plus inspection. Meantime, game sessions from the smallest cluster (C2) do not have any evident strategies and mostly have a small number of actions. Therefore, we have made an assumption that these game sessions are outlying, where students' in-game behavior differs from the in-game behavior of the majority of students.

Outlier detection has confirmed that game sessions from cluster C2 are outlying (RQ3). Among twenty-two detected outliers eighteen form entire cluster C2. Within outlier detection, we have exposed the main reasons for arising outliers: delivering care before planning and inconsistent planning. Outliers can occur due to a bad/wrong understanding of the game processor's unawareness of good practice guidelines. The game allows players to override the good practice, for example, minimize the scheduling step and jump to the delivering step. The outliers have pointed out the students, who more engaged in the gaming process and not in the learning process. They do clicks and try to progress without using previous courses, knowledge, and skills.

6 Conclusion

The paper presents an analysis of students' behavior during playing a simulation game for nurse training. Within this study, such aspects of a serious game as engagement and learning have been investigated with the implementation of statistical criteria, data visualization, n-gram models, clustering, and outlier detection.

In terms of engagement, we can conclude that students were deeply involved in the learning process due to the time that they spent and actions that they did to achieve the main in-game activities such as scheduling, reading patients' documentation, and delivering care. Furthermore, the number of game actions (average values: 252 actions for successful sessions and 148 actions for lost sessions) and game sessions per student (81 students played more than 1 time) confirm the interest and willingness to succeed in the proposed mission. However, the number of failed game sessions points the difficulty of successful mission completing in medium-difficulty scenario. To improve the user's experience and avoid frustration, we are going to orient the future developments on a safekeeping of the global game functionality. This insight will help to avoid the repetition of actions (e.g. actions related to care delivery tasks) performed in previous game sessions. The student, who plays the scenario the second time will begin the game with the previous planning and with the previous delivered care. Hence, the students' experience can be improved by adding in-game dynamic feedback which alerts on the scheduling step about hard constraint violation.

The majority of students stand by the same in-game strategy, which implies being aware of the patient's pathology, doctor's prescriptions, and level of required care. The majority of frequent patterns are devoted to scheduling and inspection. In real life this situation is usually the opposite: nurses spend most of the time delivering care and abstain the scheduling and checking the workload. As a consequence, the patient-to-nurse ratio is growing which leads to professional dissatisfaction or burns out [64, 65]. The obtained outcomes showed that the game impacts students' awareness and highlights the importance of scheduling and work organization. Even if the game was not used as an interventional tool, the game sessions have revealed awareness of nurse students about gaining information required to prepare and to plan patients' care in a medical unit. Also they have realized that their role consists of both delivering care, elaborating plans and managing a team. These fact points learning positive outcomes, however, there are players with uncertain strategies and characteristics - outliers. Outlier detection exposed the limitations in understanding the game process and lack of skills for the structuring of knowledge of some players.

Students made a different conscious effort to accomplish the educational goal. Success does not depend only on their continued perseverance. The exact reason for the game failure is not evident. The loss of the session can be explained by forgetting to deliver crucial care or by issues in following a correct global care plan. Generally, these aspects refer that students faced educational difficulties, which they were unable to overcome without any help. In our case, we believe that avoiding the delivering care step during the repetition of the game, should allow to better identify the root causes of failure. Implementing this, we will use learning analytics results to improve the gamification.

Acknowledgement

This work is a part of a global innovative IT program whose partners are University Champollion in France and the French Regional Healthcare Agency (Occitanie).

References

[1] R. Flin, R. Patey, R. Glavin, and N. Maran, "Anaesthetists' non-technical skills," *BJA: British Journal of Anaesthesia*, vol. 105, no. 1, pp. 38–44, 06 2010. doi: 10.1093/bja/aeq134

[2] D. R. Michael and S. L. Chen, *Serious games: Games that educate, train, and inform*. Muska & Lipman/Premier-Trade, 2005.

[3] M. Zyda, "From visual simulation to virtual reality to games," *Computer*, vol. 38, pp. 25 – 32, 10 2005. doi: 10.1109/MC.2005.297

[4] P. Hallinger, R. Wang, C. Chatpinyakoop, V.-T. Nguyen, and U.-P. Nguyen, "A bibliometric review of research on simulations and serious games used in educating for sustainability, 1997–2019," *Journal of Cleaner Production*, vol. 256, 2020. doi: 10.1016/j.jclepro.2020.120358

[5] G. Petri and C. Gresse von Wangenheim, "How games for computing education are evaluated? a systematic literature review," *Computers & Education*, vol. 107, pp. 68 – 90, 2017. doi: 10.1016/j.compedu.2017.01.004

[6] C. Alonso-Fernndez, A. Calvo-Morata, M. Freire, I. Martnez-Ortiz, and B. Fernndez-Manjn, "Applications of data science to game learning analytics data: A systematic literature review," *Computers & Education*, vol. 141, p. 103612, 2019. doi: 10.1016/j.compedu.2019.103612

[7] R. Wainess, H. Oneil, and E. Baker, "Classification of learning outcomes: Evidence from the computer games literature," *The Curriculum Journal*, vol. 16, pp. 455–474, 12 2005. doi: 10.1080/09585170500384529

[8] D. Clow, "An overview of learning analytics," *Teaching in Higher Education*, vol. 18, no. 6, pp. 683–695, 2013. doi: 10.1080/13562517.2013.827653 Publisher: Taylor & Francis.

[9] M. Brown, *Learning analytics: The coming third wave (EDUCAUSE Learning Initiative Brief)*. Retrieved from EDUCAUSE library https://net. educause. edu/ir/library/pdf/ELIB1101.pdf, 2011.

[10] N. Zaric, R. Roepke, M. Gottschlich, and U. Schroeder, "Interactive Gamification Analytics Tool (IGAT)," in *Games and Learning Alliance*, ser. Lecture Notes in Computer Science, I. Marfisi-Schottman, F. Bellotti, L. Hamon, and R. Klemke, Eds. Cham: Springer International Publishing, 2020. doi: 10.1007/978-3-030-63464-3_6. ISBN 978-3-030-63464-3 pp. 57–68.

[11] B. Heilbrunn, P. Herzig, and A. Schill, *Gamification analytics methods and tools for monitoring and adapting gamification designs*. Springer, 2017.

[12] C. Dichev, D. Dicheva, and K. Irwin, "Gamification driven learning analytics," in *Proceedings of the 13th International Conference on e-Learning*, 2018, pp. 70–76.

[13] D. S. Epstein, A. Zemski, J. Enticott, and C. Barton, "Tabletop Board Game Elements and Gamification Interventions for Health Behavior Change: Realist Review and Proposal of a Game Design Framework," *JMIR Serious Games*, vol. 9, no. 1, p. e23302, Mar. 2021. doi: 10.2196/23302

[14] X. Li, T. Wang, and H. Wang, "Exploring n-gram features in clickstream data for mooc learning achievement prediction," in *Database Systems for Advanced Applications*, 03 2017. doi: 10.1007/978-3-319-55705-2_26. ISBN 978-3-319-55704-5 pp. 328–339.

[15] F. Laamarti, M. Eid, and A. El Saddik, "An overview of serious games," *International Journal of Computer Games Technology*, vol. 2014, 10 2014. doi: 10.1155/2014/358152

[16] M. Liu, J. Kang, S. Liu, W. Zou, and J. Hodson, "Learning analytics as an assessment tool in serious games: A review of literature," *Serious Games and Edutainment Applications: Volume II*, pp. 537–563, 03 2017. doi: 10.1007/978-3-319-51645-5_24

[17] M. Priyaadharshini, N. Natha Mayil, R. Dakshina, S. Sandhya, and R. B. Shirley, "Learning analytics: Game-based learning for programming course in higher education," *9th World Engineering Education Forum (WEEF 2019) Proceedings : Disruptive Engineering Education for Sustainable Development*, vol. 172, pp. 468–472, 2020. doi: 10.1016/j.procs.2020.05.143

[18] A. Caldern and M. Ruiz, "A systematic literature review on serious games evaluation: An application to software project management," *Computers& Education*, vol. 87, pp. 396 – 422, 2015. doi: 10.1016/j.compedu.2015.07.011

[19] A. All, E. Castellar, N. Castellar, and J. Looy, "Measuring effectiveness in digital game-based learning: A methodological review," *International Journal of Serious Games*, vol. 1, 05 2014. doi: 10.17083/ijsg.v1i2.18

[20] F. Spyridonis and D. Daylamani-Zad, "A serious game to improve engagement with web accessibility guidelines," *Behaviour & Information Technology*, 01 2020. doi: 10.1080/0144929X.2019.1711453

[21] A. Abdul and P. Felicia, "Gameplay engagement and learning in game-based learning: A systematic review," *Review of Educational Research*, vol. 85, 03 2015. doi: 10.3102/0034654315577210

[22] D. Rodriguez-Cerezo, A. Cabezuelo, M. Gomez-Albarran, and J.-L. Sierra, "Serious games in tertiary education: A case study concerning the comprehension of basic concepts in computer language implementation courses," *Computers in Human Behavior*, vol. 31, pp. 558–570, 02 2014. doi: 10.1016/j.chb.2013.06.009

[23] A. Tlili, F. Essalmi, M. Jemni, and D. Kinshuk, "An educational game for teaching computer architecture: Evaluation using learning analytics," in *5th International Conference on Information & Communication Technology and Accessibility (ICTA)*, 12 2015. doi: 10.1109/ICTA.2015.7426881 pp. 1–6.

[24] G. Gris, H. Alves, G. Assis, and S. R. L. de Souza, "The use of adapted games for assessment of mathematics and monetary skills," 2017. doi: 10.9788/TP2017.3-12Pt

[25] H. Pope and C. Mangram, "Wuzzit trouble: The influence of a digital math game on student number sense," *International Journal of Serious Games*, vol. 2, 12 2015. doi: 10.17083/ijsg.v2i4.88

[26] N. Shin, L. Sutherland, C. Norris, and E. Soloway, "Effects of game technology on elementary student learning in mathematics," *British Journal of Educational Technology*, vol. 43, 07 2012. doi: 10.1111/j.1467-8535.2011.01197.x

[27] J. Asbell-Clarke, E. Rowe, V. Almeda, T. Edwards, E. Bardar, S. Gasca, R. Baker, and R. Scruggs, "The development of students computational thinking practices in elementary-and middle-school classes using the learning game, zoombinis," *Computers in Human Behavior*, vol. 115, p. 106587, 2021. doi: 10.1016/j.chb.2020.106587

[28] M.-T. Cheng, Y.-W. Lin, and H.-C. She, "Learning through playing virtual age: Exploring the interactions among student concept learning, gaming performance, in-game behaviors, and the use of in-game characters," *Computers & Education*, vol. 86, pp. 18 – 29, 2015. doi: 10.1016/j.compedu.2015.03.007

[29] N. Cagiltay, E. Ozcelik, and N. Ozcelik, "The effect of competition on learning in games," *Computers & Education*, vol. 87, pp. 35–41, 04 2015. doi: 10.1016/j.compedu.2015.04.001

[30] U. Tellioglu, G. Xie, J. Rohrer, and C. Prince, "Whale of a crowd: Quantifying the effectiveness of crowd-sourced serious games," in *2014 Computer Games: AI, Animation, Mobile, Multimedia, Educational and Serious Games (CGAMES)*, 07 2014. doi: 10.1109/CGames.2014.6934151 pp. 1 – 7.

[31] D. Kerr and G. K. W. K. Chung, "Identifying key features of student performance in

educational video games and simulations through cluster analysis," in *EDM 2012*, 2012. doi: 10.5281/zenodo.3554647

[32] M. Denden, A. Tlili, F. Essalmi, and M. Jemni, "Implicit modeling of learners personalities in a game-based learning environment using their gaming behaviors," *Smart Learning Environments*, vol. 5, pp. 1–19, 2018. doi: 10.1186/s40561-018-0078-6

[33] E. Rowe, M. V. Almeda, J. Asbell-Clarke, R. Scruggs, R. Baker, E. Bardar, and S. Gasca, "Assessing implicit computational thinking in zoombinis puzzle gameplay," *Computers in Human Behavior*, p. 106707, 2021. doi: 10.1016/j.chb.2021.106707

[34] D. Novoseltseva, C. P. Lelardeux, and N. Jessel, "Factors affecting success in a digital simulation game for nurse training," in *International Conference on Games and Learning Alliance*. Springer, 2020. doi: 10.1007/978-3-030-63464-32_5 pp. 263–272.

[35] H.-T. Hou, "Exploring the behavioral patterns of learners in an educational massively multiple online role-playing game (mmorpg)," *Computers & Education*, vol. 58, no. 4, pp. 1225 – 1233, 2012. doi: 10.1016/j.compedu.2011.11.015

[36] E. L. Snow, L. K. Allen, M. E. Jacovina, and D. S. McNamara, "Does agency matter?: Exploring the impact of controlled behaviors within a game-based environment," *Computers & Education*, vol. 82, pp. 378 – 392, 2015. doi: 10.1016/j.compedu.2014.12.011

[37] J. Kang and M. Liu, "Investigating navigational behavior patterns of students across at-risk categories within an open-ended serious game," *Technology, Knowledge and Learning*, pp. 1–23, 2020. doi: 10.1007/s10758-020-09462-6

[38] V. Owen and R. Baker, "Fueling prediction of player decisions: Foundations of feature engineering for optimized behavior modeling in serious games," *Technology, Knowledge and Learning*, 11 2018. doi: 10.1007/s10758-018-9393-9

[39] J. Kang, M. Liu, and W. Qu, "Using gameplay data to examine learning behavior patterns in a serious game," *Computers in Human Behavior*, vol. 72, pp. 757–770, 2017. doi: 10.1016/j.chb.2016.09.062

[40] C. Loh and S. Yanyan, "Measuring expert performance for serious games analytics: From data to insights," in *Serious Games Analytics*, 06 2015, pp. 101–134.

[41] ——, "Measuring the (dis-)similarity between expert and novice behaviors as serious games analytics." *Education and Information Technologies*, vol. 20, pp. 5–19, 03 2015. doi: 10.1007/s10639-013-9263-y

[42] J. Kang, D. An, L. Yan, and M. Liu, "Collaborative problem-solving process in a science serious game: Exploring group action similarity trajectory," *Proceedings of the 12th International Conference on Educational Data Mining*, 2019.

[43] A. Slimani, E. Fatiha, E. Lotfi, O. Bakkali Yedri, and M. Sbert, "Learning analytics through serious games: Data mining algorithms for performance measurement and improvement purposes," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 31, p. 46, 01 2018. doi: 10.3991/ijet.v13i01.7518

[44] M. Martinez-Garza and D. Clark, "Investigating epistemic stances in game play with data mining," *International Journal of Gaming and Computer-Mediated Simulations*, vol. 9, pp. 1–40, 07 2017. doi: 10.4018/IJGCMS.2017070101

[45] G. Chung, "Guidelines for the design and implementation of game telemetry for serious games analytics," in *Serious games analytics*, 01 2015, pp. 59–79.

[46] M. Liu, J. Lee, J. Kang, and S. Liu, "What we can learn from the data: A multiple-case study examining behavior patterns by students with different characteristics in using a serious game," *Technology, Knowledge and Learning*, vol. 21, no. 1, pp. 33–57, 2016. doi: 10.1007/s10758-015-9263-7

[47] V. J. Hodge and J. Austin, "A survey of outlier detection methodologies," *Artificial Intelligence Review*, vol. 22, no. 2, pp. 85–126, Oct 2004. doi: 10.1007/s10462-004-4304-y

[48] E. M. Knorr, R. T. Ng, and V. Tucakov, "Distance-based outliers: Algorithms and applications," *The VLDB Journal*, vol. 8, no. 3-4, pp. 237–253, Feb. 2000. doi: 10.1007/s007780050006

[49] H. Aldowah, H. Al-Samarraie, and W. Fauzy, "Educational data mining and learning analytics for 21st century higher education: A review and synthesis," *Telematics and Informatics*, vol. 37, pp. 13–49, 4 2019. doi: 10.1016/j.tele.2019.01.007

[50] N. Riem, S. Boet, M. Bould, W. Tavares, and V. Naik, "Do technical skills correlate with non-technical skills in crisis resource management: a simulation study," *British journal of anaesthesia*, vol. 109, no. 5, pp. 723–728, 2012. doi: 10.1093/bja/aes256

[51] S. Gough, M. Hellaby, N. Jones, and R. MacKinnon, "A review of undergraduate interprofessional simulation-based education (ipse)," *Collegian*, vol. 19, no. 3, pp. 153–170, 2012. doi: 10.1016/j.colegn.2012.04.004

[52] M. Cornes, "Book review of interprofessional education in the united kingdom (1997-2013), by hugh barr , marion helme , and lynda d'avray , fareham : Caipe , 2014 , 131 pages," *Journal of interprofessional care*, vol. 29, p. 1, 11 2014. doi: 10.3109/13561820.2014.981479

[53] C. E. Reese, P. R. Jeffries, and S. A. Engum, "Learning together: Using simulations to develop nursing and medical student collaboration," *Nursing education perspectives*, vol. 31, no. 1, pp. 33–37, 2010.

[54] C. P. Lelardeux, D. Panzoli, V. Lubrano, V. Minville, P. Lagarrigue, and J.-P. Jessel, "Training the operating room staff in a virtual multiplayer and real-time environment to prevent adverse events: study of team situation awareness and decision making using the learning game 3d virtual operating room," *5th International Conference on Healthcare systems ergonomics and Patient safety - Healthcare and society : new challenges, new opportunities (HEPS 2016)*, 2016.

[55] C. Pons Lelardeux, D. Panzoli, M. Galaup, V. Minville, V. Lubrano, P. Lagarrigue, and J.-P. Jessel, "3d real-time collaborative environment to learn teamwork and non-technical skills in the operating room," in *International Conference on Interactive Collaborative Learning*, 01 2017. doi: 10.1007/978-3-319-50337-0_12. ISBN 978-3-319-50336-3 pp. 143–157.

[56] J. Hu and L. Feijs, "A distributed multi-agent architecture in simulation based medical training," *T. Edutainment*, vol. 3, pp. 105–115, 01 2009. doi: 10.1007/978-3-642-11245-4_10

[57] Y. Xu, Y. Lau, L. J. Cheng, and S. T. Lau, "Learning experiences of game-based educational intervention in nursing students: A systematic mixed-studies review," *Nurse Education Today*, vol. 107, p. 105139, Dec. 2021. doi: 10.1016/j.nedt.2021.105139

[58] O. Petit-dit Dariel, T. Raby, F. Ravaut, and M. Rothan-Tondeur, "Developing the serious games potential in nursing education," *Nurse education today*, 01 2013. doi: 10.1016/j.nedt.2012.12.014

[59] C. Kilmon, L. Brown, S. Ghosh, and A. Mikitiuk, "Immersive virtual reality simulations in nursing education," *Nursing education perspectives*, vol. 31, pp. 314–7, 11 2009.

[60] C. Pons Lelardeux, H. Pingaud, M. Galaup, A. Ramolet, and P. Lagarrigue, "The challenge of designing interactive scenarios to train nurses on rostering problems in a virtual clinical unit," *Advances in Intelligent Systems and Computing*, vol. 1, 09 2018. doi: /10.1007/978-3-030-11932-4_56

[61] C. Loh and S. Yanyan, "Maximum similarity index (msi): A metric to differentiate the performance of novices vs. multiple-experts in serious games," *Computers in Human Behavior*, vol. 39, pp. 322–330, 10 2014. doi: 10.1016/j.chb.2014.07.022

[62] T. Martin, C. Smith, N. Forsgren, A. Aghababyan, P. Janisiewicz, and S. Baker, "Learning fractions by splitting: Using learning analytics to illuminate the development of

mathematical understanding," *Journal of the Learning Sciences*, vol. 24, 08 2015. doi: 10.1080/10508406.2015.1078244

[63] S. Ramaswamy, R. Rastogi, and K. Shim, "Efficient algorithms for mining outliers from large data sets," in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 2000. doi: 10.1145/342009.335437 pp. 427–438.

[64] L. Aiken, S. Clarke, D. Sloane, J. Sochalski, and J. Silber, "Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction," *JAMA : the journal of the American Medical Association*, vol. 288, pp. 1987–93, 11 2001. doi: 10.1001/jama.288.16.1987

[65] M. D. McHugh, A. Kutney-Lee, J. P. Cimiotti, D. M. Sloane, and L. H. Aiken, "Nurses' widespread job dissatisfaction, burnout, and frustration with health benefits signal problems for patient care," *Health Affairs*, vol. 30, no. 2, pp. 202–210, 2011. doi: 10.1377/hlthaff.2010.0100