Calibration of Game Dynamics for a More Even Multi–Player Experience

Mauro Scanagatta
mscanagatta@fbk.eu
Fondazione Bruno Kessler
Trento, Italy

Michela Ferron
ferron@fbk.eu
Fondazione Bruno Kessler
Trento, Italy

Giulia Deppieri
gdeppieri@fbk.eu
Fondazione Bruno Kessler
Trento, Italy

Annapaola Marconi
marconi@fbk.eu
Fondazione Bruno Kessler
Trento, Italy

Abstract

When designing a serious game it is paramount to take into account that the players' performance may greatly vary. In many types of games the performance will impact the game experience, as it is related to a point acquisition mechanic. In specific contexts, such as multi-player non-competitive or collaborative games, the designer may desire to minimize this variability, reducing the difference in game progression between low-performing and high-performing players. In this paper we detail such a context: an educational serious game for sustainability addressed to classes of a primary school, where the low performance of a class may not be necessarily related to the children's effort. We thus present and evaluate a novel approach for the calibration of the game dynamics for achieving a more even player experience.

CCS Concepts

• **Human-centered computing → Human computer interaction (HCI)**; • **Computing methodologies → Artificial intelligence**.

Keywords

Education and learning-related technologies; User-Adaptive interaction and personalization

ACM Reference Format:

Mauro Scanagatta, Michela Ferron, Giulia Deppieri, and Annapaola Marconi. 2020. Calibration of Game Dynamics for a More Even Multi–Player Experience. In *25th International Conference on Intelligent User Interfaces (IUI '20), March 17–20, 2020, Cagliari, Italy*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3377325.3377521

1 Introduction

Serious games can be defined as games "in which education, in its various forms, is the primary goal, rather than entertainment" [35]. This research topic is growing, and serious games are routinely applied for learning, education and training [13, 16, 20, 22, 29].

In order to fully exploit serious games' benefits, players need to remain active in the game over prolonged periods of time. Player's retention is particularly relevant in games with a purpose, where keeping the player engaged for longer may result in a more permanent and significant behavior change [1, 19, 21, 43]. This requires for sure a careful game design, to meet the player preferences, type and play style, and to enable an adequate player experience. In addition to this, a key factor for players' retention is game adaptivity and personalization [39]: that is, being able to adjust the content and interaction schemes of games, as virtual environments, to the knowledge level, skill, preferences and experience of the users. This is a well recognized and experimented practice in entertainment games, where dynamic difficulty adjustment (DDA) [10, 11, 18] and procedural content generation (PCG) [28, 42, 45] techniques are wildly investigated and applied.

Recent studies conducted on the effects of serious games show that also in this domain, personalization and adaptivity can promote user acceptance, motivation and retention [16, 20, 22, 25, 30, 32, 33, 41]. Despite these findings, the field of adaptive serious games is still largely unexplored, and has been gaining attention from researchers only in recent years [23, 39].

In this paper we focus on a serious game, *WEEE R Robots*, whose purpose is to promote responsible consumption and disposal of electric and electronic equipment (EEE), targeting children in primary schools, as a way to decrease the impact on the environment. *WEEE R Robots* intends to increase the awareness on using electric and electronic equipment for longer (reduce), repairing and reusing still functioning electric items (reuse), and disposing them in a proper way (recycle).

The targeted problem has a significant sociological and environmental importance, as the European Commission reported that the waste of electrical and electronic equipment (WEEE), with about 9 million tonnes generated in 2005, which are expected to grow to more than 12 million tonnes by 2020, can be considered the fastest waste–growing stream in the EU [6]. On the other hand, the reuse of EEE and the recycling of WEEE offer substantial opportunities in terms of extension of End of Life (EoL) of appliances and of making secondary raw materials available on the market.

WEEE R Robots is an education for sustainability game that involves the whole school (teachers, children, and their families) for

IUI '20, March 17–20, 2020, Cagliari, Italy

ACM ISBN 978-1-4503-7118-6/20/03…$15.00
https://doi.org/10.1145/3377325.3377521

6 weeks. During this time, each class of children would compete as a single player, obtaining game points as a reward for bringing from home unused or broken electric and electronic objects, that are either to be recycled or to be reused, and by completing surveys on the reduction topic. In the game the children can then spend the game points buying upgrades of a virtual robot, acting as an extrinsic reward for their efforts. The school as a whole has the objective of building a team of robots that fights against environmental pollution.

The challenge we faced is the *calibration* of the game, in terms of cost of the various robot's parts, to ensure an even player experience between the classes. In *WEEE R Robots* this is particularly important for two main reasons. First of all, this calibration allows to emphasize the cooperative aspects of the game, building together, as a school, a team of virtual robots, rather than the competitive ones (i.e., classe competing with each other to build the 'best' robot). Moreover, the classes' performance may greatly vary, but the low performance of a class is not necessarily related to the children's commitment and effort. Children might not bring to school a high number of electronic items because their families do not have time or are not willing to support them in the campaign, or because they are already in-line with the sustainability objectives pursued by the game (i.e. limited consumption, correct reuse and disposal of EEE), or since children are coming from low-income families, thus having less objects to bring to the class.

Game content adaptivity and personalization in a multi–player setting, which is particularly relevant for non-competitive and collaborative games, is a well recognized and still uninvestigated problem [39]. In this paper, we propose and evaluate a novel approach for calibrating game dynamics in a multi–player educational game, which aims at providing a more even experience to players.

In Section 2 we first present relevant work in the literature. In Section 3 we detail the context of the gamified system. In Section 4 we explain the calibration that was performed in the first iteration of the project, using a naïve approach. In Section 5 we present a novel approach, that is probably able to better calibrate the game dynamics, and we discuss the evaluation results. Section 6 concludes our paper.

2 Related work

Serious games have proven to be successful to raise awareness, increase participation and promote sustainable behaviour in social, environmental and health domains [2, 12, 13, 37]. Games with a purpose leverage on rewarding experience and fun to raise awareness and increase people's motivation to change their behavior consistently to certain values. Environmental sustainability is one of the domains where gamification has been successfully employed, with applications ranging from sustainable mobility [13, 29], tourism [31], motivating informed action about climate change [24], energy saving in the office [34] and in family environments [38].

A current limitation of serious games is that they often fail to keep the player engaged in the long–term [15]. That is especially unfortunate, because a behavioral change cannot be immediately internalized; rather, it must be reinforced long enough to successfully form new habits [44].

According to the flow theory by Csikszentmihalyi [8], one of the key enabling factors for motivation is the calibration of challenges and skills. Flow theory has been connected both to games [5, 9] and also explicitly to serious games [36]. Thus, an important aspect for player retention is being able to minimize players' frustration or boredom calibrating the difficulty of the game to the player's skills and performance [39].

In entertainment games DDA techniques are wildly exploited to adapt the level of difficulty so that it is tailored to the players' abilities [10, 11, 18]. DDA has also been implemented in the context of serious games, although in this context its application is still young and growing, especially for what concerns dynamic adjustment [40].

For instance, in [4] the authors employed a gradually increasing challenge difficulty (both in terms of time allotted and conceptual difficulty) in order to encourage the user to experience an app tackling the problem of low adult literacy. The authors reported that the increasing difficulty of the challenges contrasted feelings of boredom in participants. Another example is Jagust et al.'s study [20], in which two dynamic adaptation scenarios, time-based and score-based, were evaluated within a system supporting young adults math learning.

DDA was also employed by Howell and Veale (2006) [17] within a game-based ITS for learning linguistic abilities. This system attempted to keep learners immersed in the game, in a state of flow, by dynamically adjusting the level of difficulty, while at the same time respecting the constraints to achieve the learning targets. Another example is the study from Kickmeier-Rust et al. on adaptive educational games [22], where the customization was in the types of badges presented and in the feedback given to players. The authors emphasized the importance of developing an educational game which dynamically balanced game dynamics to achieve superior gaming experience and educational gains.

Additional challenges arise when the adaptation of game dynamics needs to be applied in multi–player games, which is still an uninvestigated field [40]. This is true for entertainment games in general, and serious games in particular. Adaptive algorithms, in this context, need to properly calibrate the game to meet the expectations and skills of several players, that might have very heterogeneous performances and preferences [40]. Moreover, automatic game adjustment, if not carefully designed and properly implemented, can have a strong impact on the balancing and the perception of the balancing by the players, which in turn may interfere with their game experience [14]. Particular care needs to be made in handling the game adaptation in such a way that it is transparent to players. Gerling et al. compared visible and transparent adjustment strategies in an established motion-based game [14]. The study revealed that explicitly observable calibrations might negatively impact players' self-esteem and feelings of relatedness in player pairs, whereas transparent adjustments seemed to improve self-esteem and reduce performance differences without necessarily affecting the player experience.

3 WEEE R Robots

The main objective of *WEEE R Robots* is to induce a short term commitment to sustainable practices that fosters consistent behavior

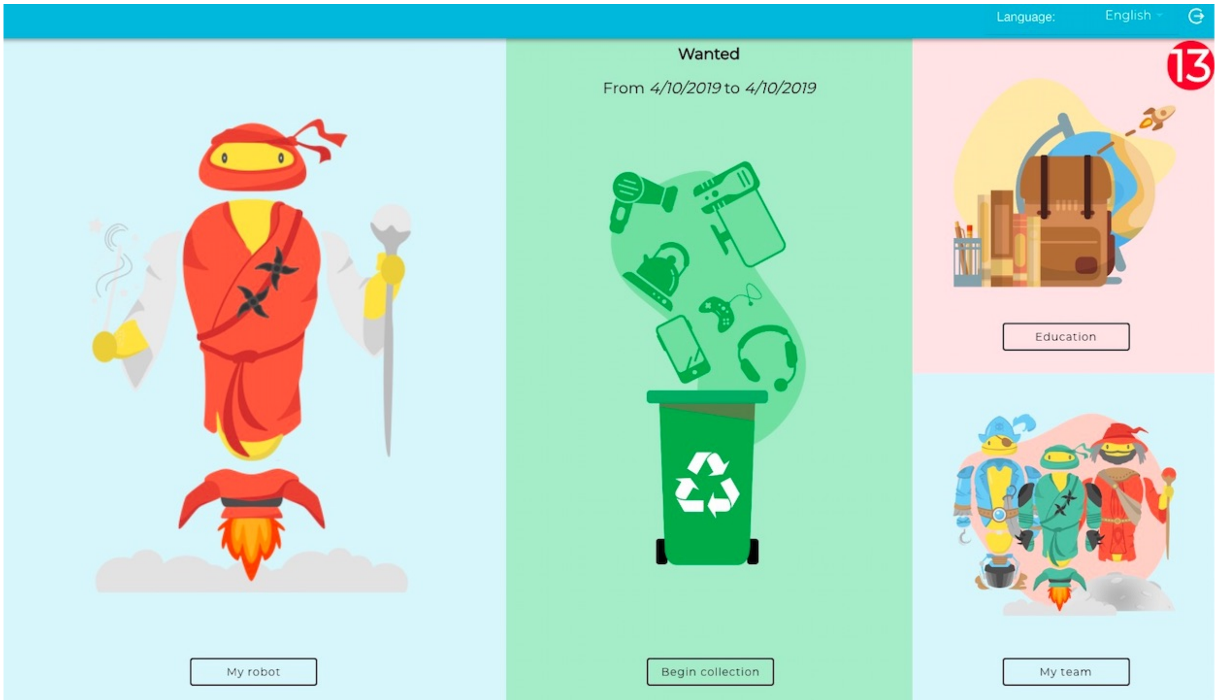

Figure 1: *WEEE R Robots* web app: Homepage.

change in the long term. *WEEE R Robots* is designed as a collaborative playful experience conveyed at primary schools and tailored to children aged 6 to 10. In this Section we describe the design of *WEEE R Robots* and its first implementation in a primary school of Trento (Italy), outlining the design rationales that guided its definition.

3.1 An overview of *WEEE R Robots* campaign

WEEE R Robots campaigns are structured as school challenges exploiting a gameful education approach to promote responsible consumption of electric and electronic equipment (EEE), by increasing children's awareness and good practices regarding the 3Rs (reduce, reuse, and recycle): using still working electric and electronic equipment as long as possible (reduce), repairing broken electronic items or giving them to others if they are no more needed (reuse), and disposing them correctly if they cannot be repaired (recycle), as a way to decrease the impact on the environment.

The game is organized over 6 school weeks, each one characterized by a thematic challenge in which pupils and their families collect small electric and electronic items belonging to a specific category (e.g., small kitchen appliances, electric and electronic games, smartphones and tablets). Every week, pupils and their families search at home for electric appliances no longer in use or broken. For each object, they fill in a card with information useful to determine the object status and destination (e.g., age, whether it is broken or damaged). At home, pupils also answer one weekly question about their waste reduction habits (e.g., how old is your fridge?). This question is meant to raise their awareness of responsible consumption (reduce) and contributes to the in-game rewards.

On a predefined day of the week, children bring their objects and cards to school. Together with their teacher, they use the *WEEE R Robots* web app (see Figure 1 for the web app homepage) to classify each object with the information written in the item card. With this information, the web app provides information about the residual life of the electronic item and on its destination, in terms of the bin or box to be placed in. Reusable items are placed in the EEE reuse box, repairable items in the *Green Idea Technologies'* Ecobox, and waste items in the WEEE recycle bin.

At the end of the collection process, the web app shows the outcomes of the collection (see Figure 2) in terms of: i) gained credits by the class (1 *Reuse credit* for each repairable or reusable item of EEE, 1 *Recycle credit* for each item of WEEE, 1 *Reduce credit* for each year of the item they have in use at home); ii) type and amount of materials (e.g., iron, plastic, copper, glass, gold, silver) obtained thanks to recycling; and iii) total items' weight and CO2-equivalent savings.

The performances of the weekly challenge are conveyed through the metaphor of the robot. Each robot is composed of five body parts (head, left arm, right arm, legs, and chest). When children have collected enough credits, they can unlock body parts of the robot: each element varies in terms of appearance (colour, shape, details). Children can choose which element to purchase as well as a particular version of that element among a given selection, making each robot unique. Once a robot part is purchased, new upgraded versions of the same body part are unlocked and can be purchased by the children. There are in total three levels of upgrade, for each of the five parts. The game aim is to build a team of virtual robots (one robot by each class) that will fight environmental pollution.

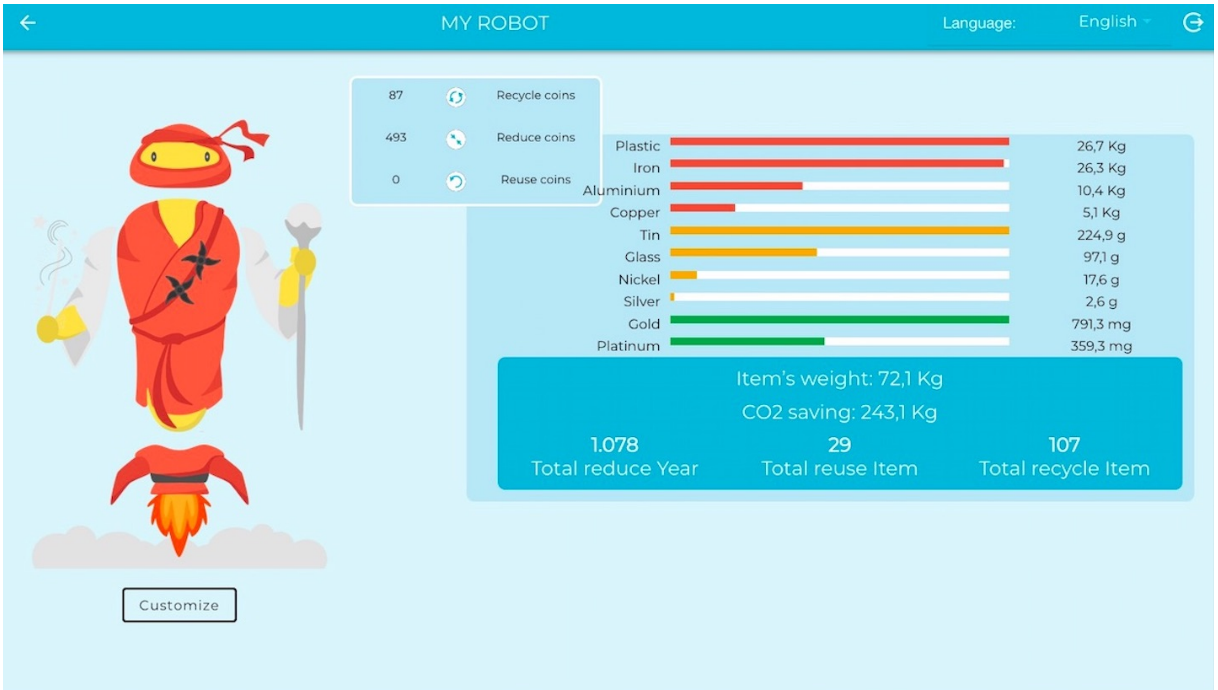

Figure 2: *WEEE R Robots*: Class-level results and robot.

Children can also inspect the overall impact of the campaign for the whole school and the robots built by pupils in the other classrooms (see Figure 3), and access educational content on environmental sustainability.

3.2 Design rationales

Bringing gameful activities into the school leads to an increased appeal in the young generations and helps them to engage in practices with an educational function.

The design of *WEEE R Robots* refers to the following design rationales:

#1 Together we make a difference. The game metaphors are designed to convey the message that individual actions result in a larger impact. The game aim is to build a team of virtual robots, made of recycled materials and reconditioned items, that will fight environmental pollution. Children collaborate to create their own robot in the class. Through the Web App, pupils can also view the overall impact of the campaign at school level and the team of robots built by the classes (see Figure 3).

#2 Education for sustainability. The game is designed to teach children to identify key features of EEE objects, to classify them, and to decide of their "fate" based on that classification. For example, if an object is still functioning, it can be reused or donated. If it is damaged beyond repair, it must be correctly disposed and recycled. In *WEEE R Robots*, objects that have a residual value and could be repaired are conveyed to *Green Idea Technologies* (a local company) to be refurbished and resold, and part of the revenues is returned to the school for purchasing refurbished stationery. In this way, pupils learn the correct disposal of EEE items along with the value of repairing and reusing.

#3 Raising awareness. WEEE R Robots Web App promotes the values of the 3Rs (Reduce, Reuse, Recycle), induces a change in knowledge, awareness and attitude towards WEEE/EEE. At the end of the weekly collection, the app shows, in addition to the credits earned, information about the impact of the collection in terms of type and amount of material obtained thanks to recycling and CO2-equivalent saving.

#4 Integration with teaching activities. The game presents educational content that can be easily integrated in the classroom activities. Through the Web App, children can exploit new educational material that unlocks after each weekly collection challenge. The educational multimedia material (e.g. slide sets, songs, games and videos) explains the value of recycling and the problems connected to WEEE disposal and recycling.

#5 Involving families. We leverage on children's engagement to involve the whole family in the collection campaign. Families are involved throughout the campaign duration and their role is fundamental for the collection and classification of electronic items at home. In this way, we aim at inducing a larger change in the whole family's attitudes and habits, for a long-term effect and a larger social reach.

#6 A collective and collaborative game. The campaign allows the class to buy elements of the robot regardless of how many items the single pupil brings to school. The robot customization is a

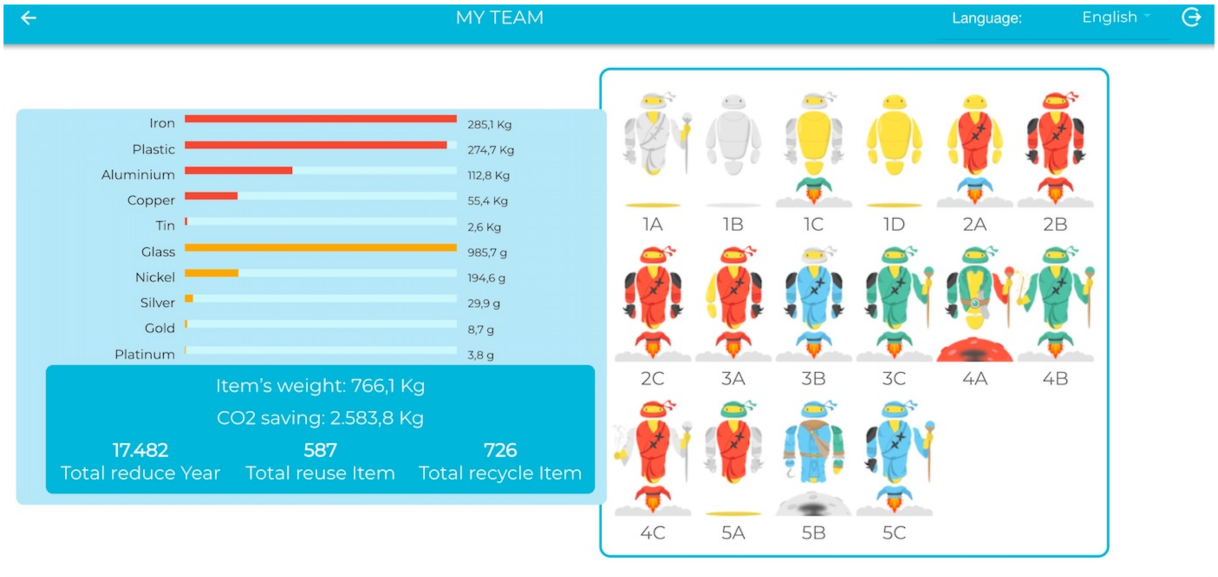

Figure 3: *WEEE R Robots* web app: School-level results and team of robots.

participatory activity in which the class has to collaborate to buy robot upgrades.

4 First trial and the calibration problem

WEEE R Robots was first implemented and evaluated within a primary school in Trento (Italy) in the spring of 2019 (8 April - June 2, 2019). 345 students from 16 classes, aged from 6 to 10 years, participated in the game and managed to collect more than 1300 items by the end of the campaign (587 EEE and 726 WEEE).

In this first implementation, we faced a challenge: ideally, we wanted each class of children to fully experience the game and to be able to *complete and personalize* their robot by the end of the game. The robot was composed of five parts (head, two arms, legs, chest), each one customizable across three aesthetically different options. In total, each player (i.e., each class of children taken as a whole) could buy up to 15 *upgrades*, which served as the main extrinsic motivational elements to foster game participation during the 6 school weeks of the campaign. We aimed to guarantee an enjoyable game experience to both high-performance and low-performance players. In the former case, it meant that no high-performance player should have been able to fully complete their robots before the final week, leaving them with no further extrinsic motivation to continue the game. In the latter case, it meant that all players should have been able to buy a minimum set of robot's parts (completing both the first and second level of upgrades, totaling 10 parts over 15 available).

In the particular context of *WEEE R Robots*, guaranteeing an even experience across players is particularly important because the low performance of a class may not necessarily be related to a low effort or participation from the children. In *WEEE R Robots*, the acquisition of game points largely depends on the amount of EEE and WEEE brought to school. However, this is influenced by several factors that can go beyond children's commitment, such as the time and effort that their parents can, or are willing to, dedicate to the campaign, or the number of electronic items the families have at home.

In *WEEE R Robots*, our approach was to have a *decreasing* cost of the upgrades over time. In this way we aimed to slightly penalize high-performing classes, as they would buy upgrades earlier, thus paying them more. We expected this to reduce the variance of the final robot completion across players. The rate of point acquisition, intended as the number of points rewarded for each in–game action, was instead kept fixed. In this approach, we assumed that the children would not pay attention to the reduction of upgrades cost over time, thus implementing a transparent adjustment. Indeed, no children or teacher mentioned this aspect in the final assessment (see the qualitative evaluation at the end of the section).

4.1 First calibration of game dynamics

For the first trial of the campaign, there was no data available for predicting the collection performances and the corresponding rate of game points acquisition by the players, so we grounded our reasoning on expert opinion. What we did know was the size of the classes involved in the project, composed on average by 21.3 children (SD = 1.92). We estimated, for each week and each point type, the contribution provided by a typical child in our community (Table 1a).

From the estimated distribution, we extracted the predicted point acquisition for three types of class: low-performance (1st decile), normal (5th decile) and high-performance (9th decile) class. The rationale was to directly compare the effects of the upgrades cost calibration across the three groups, keeping in mind that the main goal of the project was to ensure that low-performing classes would still be able to buy a minimum number of upgrades (i.e., 10 over 15), and high-performing ones would not be able to complete the robot before the last week.

	Mean	SD
Reduce	558.7	272.14
Reuse	63.64	31.45
Recycle	31.82	23.45

(a) Estimated point acquisition statistics.

	Mean	SD
Reduce	792.56	355.21
Reuse	36.69	27.28
Recycle	45.38	24.27

(b) Recorded point acquisition statistics (summary).

Table 1: Mean and standard deviation of the point acquisition statistics.

Point type		First week	Second week	Third week	Fourth week	Fifth week	Sixth week
Reduce	Mean	157.94	131.81	159.88	80.94	126.69	135.31
	SD	31.63	86.72	132.35	69.53	66.30	96.79
Reuse	Mean	4.44	5.44	6.44	2.19	6.44	11.75
	SD	3.22	5.98	6.96	1.72	5.76	11.38
Recycle	Mean	2.69	14.13	6.25	2.63	5.50	14.19
	SD	1.58	8.91	5.00	2.19	3.54	12.51

Table 2: Recorded point acquisition statistics (detail).

From this initial estimation the cost of the robot's upgrade were determined by the authors of this paper, with the following constraints:

- In a given week, the cost of an upgrade cannot be lower than the cost of its ancestor;
- The cost of an upgrade in a given week cannot be lower that the cost of the same upgrade on the following week.

The point acquisition rate was modeled as a Dirichlet distribution, and the distribution was updated each week through Bayesian inference taking into account both our prior belief and observed rate until that moment. The cost of the robot's upgrades were manually updated according to the updated belief.

The statistics recorded at the end of the project can be found at Table 1b (summary) and Table 2 (detail). Here, the statistics regarding the *Reduce* point acquisition refer to the children's answers to the questions on their waste reduction habits, while the data regarding the *Reuse* and *Recycle* point acquisition refer to the number of EEE and WEEE items that children brought to school. If we compare the observed point acquisition rate with the predicted one (Tables 1a and 1b), we can clearly observe that we initially underestimated the Reduce and Recycle rates, while we overestimated the Reuse one. Regarding the detailed point acquisition statistics reported in Table 2, we can observe that the point acquisition varied across weeks, probably depending on the type of objects collected each week. For example, the second week was dedicated to the collection of mobile phones, smartphones, tablet PCs and laptops, and we can observe that classes earned on average 14.13 points by collecting a significant number of old smartphones and mobile phones. A particular case is the sixth and last week, which had been originally designed to collect electric and electronic games, but was instead opened in the end to the other categories because children, despite not having many electric and electronic games at home, had found in the meantime other EEE and WEEE items that could be useful to upgrade the robots.

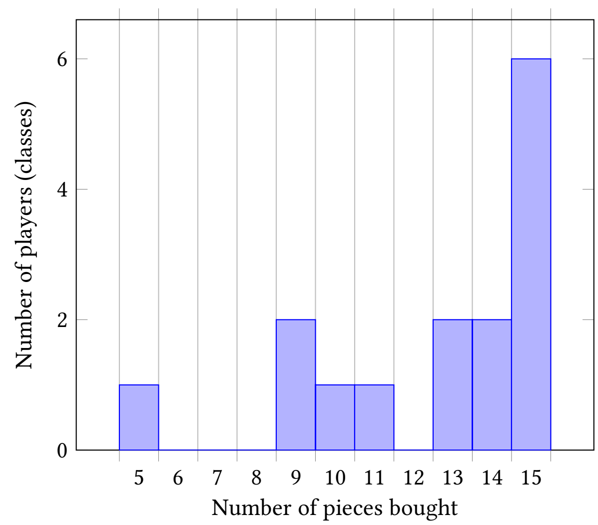

Figure 4: Distribution of the final number of upgrades bought in the first iteration of the project.

4.2 Quantitative and qualitative evaluation of the first trial

To evaluate *WEEE R Robots* first trial, we performed a quantitative analysis of children's purchasing behavior, with regard to the pieces of the robot, and a qualitative evaluation of the experience at school, with both children and teachers.

We considered only those players that were active (i.e., performed at least one game action) during the whole 6–week period of the game. 15 over 16 classes satisfied this constraint: only one class was thus excluded, as it did not obtained enough points to buy at least one upgrade, as it can be observed from Figure 3.

The mean number of pieces bought by the players was 12.06 (over a maximum of 15; SD = 3.51). Four classes out of 15 did not buy a minimum of 10 pieces (26.6%), which violated the objective we had initially set regarding low-performing players (letting all players purchase at least 10 upgrades). Conversely, no class was able to completed the robot before the end of the game, achieving our initial goal regarding high-performing players. Finally, 6 classes out of 15 (40%) completed the robot by the end of the game, obtaining a full game experience. We report the distribution of pieces bought at the end of the game in Figure 4.

As we can see in the figure, the manual calibration of the game dynamics achieved only partially the proposed objectives. In particular, we can observe that the first grade classes had the lowest performance in the school. This could be due to several reasons, ranging from potential difficulties to clearly communicate the game-related tasks to the youngest children, to possible problems in reaching the families. To tackle this issue, different solutions can be implemented that involve the planning of the campaign, for example adopting a different communication strategy with the youngest children, or the game mechanics. With regard to the game mechanics, we propose to perform the cost calibration in a more principled manner, one that would better guarantee the fulfilment of our objectives. We detail it in the next section.

	1 lvl	2 lvl	3 lvl
Mean	3.94	3.63	3.06
SD	1.73	1.75	2.08

Table 3: Statistics on piece acquisition grouped by piece level.

	Head	Chest	ArmL	ArmR	Legs
Mean	2.53	2.67	2.33	2.47	2.53
SD	0.74	0.72	0.82	0.74	0.83

Table 4: Statistics on piece acquisition grouped by piece type.

We additionally report the statistics on the pieces bought divided by piece level in Table 3, and divided by piece type in Table 4. We can observe that robot pieces of a lower level were more frequently bought, probably due do their lower cost. On the contrary, there were no clear preferences between the different piece types.

To gain a better understanding of the overall experience of *WEEE R Robots*, we conducted an evaluation at school with 15 teachers (F=15) and four classes (3 classes in the 5th grade and 1 class in the 4th grade; N=68; F=30). For the assessment with the teachers, we opted for an unstructured evaluation to explore the multidimensional complexity of the campaign and let key experiential aspects emerge directly from the teachers. With the pupils, we used a more structured evaluation on selected aspects combining questionnaires and open questions but allowed about 15 minutes for spontaneous feedback from the children. In this paper, we focus on the unstructured feedback we received from teachers and children regarding the construction of the robots.

With teachers, we adapted the customer journey mapping technique to assess the experience of services [7, 26]. After collecting informed consent forms for audio and video registration, we divided the teachers in groups of five with one facilitator per group, then asked them to individually identify the actions they carried out from the beginning to the end of the campaign. Then, we grouped individual actions on a timeline representing the duration of the campaign: this activity allowed us to identify the most significant types of actions (i.e., explaining the process to the children, classifying the objects in class, building the robot). For each group of actions, we explored three main dimensions: the emotions associated to those actions, the needs, and children's reaction. The overall evaluation with teachers took about 2 hours. To collect spontaneous feedback from the children, we asked what they liked the most and the least in the campaign. For privacy reasons, we did not record the session with children but took notes and observations. The evaluation with the children lasted about 15 minutes.

We reviewed the videos of the workshop with teachers and transcribed group discussions, using thematic analysis to iteratively elaborate on key themes [3]. Two researchers independently coded the data, developing an initial coding scheme that was iteratively discussed until a consensus was reached. Then, we categorized the data into subgroups semantically consistent and identified broader themes. In this paper, we report on the results related to the teachers' and the pupils' experience in building the robots.

- *Need for clarity on credit assignment and cost calibration.* There was sometimes confusion about how credits were assigned in relation to the number of objects to be recycled or reused, and the children's answers to the reduce questions. In particular, the children of one of the three 5th grade classes asked to add a table in the web app to clarify the number of credits assigned for each of the three good practices on recycling, reusing and reducing. The class also pointed out that the elements of the robots should cost less, to increase their own buying power. One teacher (PP5) reported: "The children were excited about building the robot, but also disappointed because the credits were never enough". Although only one teacher out of 15 mentioned this point, the request highlights the importance of a careful cost calibration to let players advance the game and receive appropriate reward.
- *Opportunity to donate objects.* During the campaign, children were used to look at the robots built by the other classes in the school. Although teachers reported competition between the classes (see below), the children were sorry when they saw that their schoolmates' robots were incomplete and showed altruistic behavior to help their friends. One teacher (GP4) said: "We thought we could lend them some of our objects, to let them earn credits".
- *Opportunity to donate credits.* This kind of altruistic behavior manifested also through the desire to donate credits. Consistently to what was reported by three of the four classes we interviewed, one teacher (GP1) said: "The children were sorry that we had many credits left but we couldn't give them to the other classes". The pupils of the 4th grade class said they had brought a great number of objects but waited until the end of the campaign to purchase the elements of the robot, so it was not until the end of the game that they figured out they had many credits left. They reported that,

if they had known how many credits they had, they would have gladly donated some to the other classes.
- *Competition.* Altruistic practices and attitudes also interwove with competitive behavior. While the children of the 4th grade class were open to donate credits, especially if they had brothers or sisters in the other classes, not all the children were happy with that. One teacher (MP4) said: "It was a competition between classes, I had two 4th grade classes and they looked at who had more credits to finish the robot first". Another teacher (MP1) reported: "That class finished the robot early, but they still brought objects, I said 'yes, so we'll earn more credits', but a couple of children looked at the robot of the other class they were competing with, and they told me I could give their objects to the other class, so I entered them in the other class".

The quantitative and the qualitative evaluation we performed highlighted the need for an improved cost calibration of game dynamics. In particular, the results of the quantitative evaluation showed only a partial fulfillment of the objectives initially set, and an uneven performance across the classes (with lower outcomes for the first grades). The qualitative assessment drew attention to the children's perception of this uneven performance, which determined their request to donate points or objects to the other classes. For these reasons, we propose an improved cost calibration algorithm, which we present in the following Section.

5 Improved calibration

In this Section we present our novel approach for the calibration of the game dynamics, one that could be better guarantee the fulfilment of our game design objectives.

Our goal is to perform the robot's pieces cost calibration in a way that ensures the following multiple goals:

- Maximize the uniformity of players' experience in the game, in terms of number of robot's pieces bought;
- Maximize the uniformity of the rate of robot's piece acquisition across different types (no type of pieces privileged over others) and weeks (uniform rate of robot completion).

As before, we put the following constraints:

- On the same week, the cost of an upgrade can't be lower than the cost of the ancestor;
- The cost of an upgrade in a given week can't be lower that the cost of the same upgrade on the following week.

We modeled the problem as a tabu-search exploration, providing as initial solution the piece acquisition prices determined in our previous experimentation.

In each iteration of the algorithm, the piece acquisition prices were randomly modified by either increasing or decreasing them by one unit, following the constraint previously outlined, in order to produce a neighbouring solution.

Using the new piece acquisition prices, we simulated a player (a class of children) as a finite state machine, shown in Figure 5, that can be summarized as:

- It has eighth states; an initial state q_a; six states $q_b, .., q_f$, each corresponding to a game week; and an end state q_e.
- Each of the state corresponding to one of the game weeks had a first action leading to that state, called rp_i (where i is the week number), simulating the point acquisition performed by the children bringing electronic devices to the school. Each class received a number of points sampled from the point acquisition distribution outlined in Table 2.
- The same state had a second action called bu_i, simulating the action of buying an upgrade using the available game points, choosing it randomly from the ones that were currently purchasable;
- When it is not possible to perform action bu_i (no more purchasable pieces, given the residual points), it would transition to the next state.

We simulated a huge number of players (10.000), assuming the same point acquisition distribution observed during the first trial, and recording their acquisition history.

In order to compare the solutions, we computed the *Between Group Variation*, as a way to estimate the variation due to the interaction between the groups. It is denoted as $SS(B)$ for Sum of Squares Between groups. It is defined as:

$$SS(B) = \sum n(\bar{x} - \bar{x}_{GM})^2 \quad (1)$$

where $\bar{x}_{GM}$ is the Grand Mean. If the sample means are close to each other (and therefore the Grand Mean) the $SS(B)$ will be small.

We took the $SS(B)$ between the number of pieces bought by each simulated players, divided by week and type, as our objective function to be minimized. A low value for this function thus indicated an higher uniformity of game experience between players, and between both weeks and piece types. We indeed wanted to avoid invalid solutions that would technically lower the variance between player's experience, such as calibrating the pieces' cost in a way that would prevent players from buying pieces during the first five weeks, and allowing to buy all of the pieces with a trivial cost in the final week.

If the value of the objective function of the neighbour solution was lower than the best one recorded, the algorithm then chose it as the current best solution. The exploration continued until no more improving solutions could be found.

We additionally recorded, for each iteration, the percentage of players that achieved to buy all the robot's pieces **before** the final week. The solution was discarded if this percentage was higher than a threshold value (0.02). This was done in order to constrain the probability of high-performance's players completing the robot before the end of the game, leaving them with nothing more to do in terms of robot personalization.

The best solution was found after four days of execution of the algorithm, when no more better solutions were found. The algorithm was executed on a laptop with four cores and 16 GB of RAM. The best solution we found achieved $SS(B)$ between the number of pieces bought, divided by week and type, of 4.94. We report the mean rates of pieces purchase, divided by type and week, in Table 5. It shows that the rate of pieces purchase was uniform between both pieces types and game weeks.

The mean number of total pieces bought by the players was 12.96, with SD 1.96 (compared to the mean 12.06 and SD 3.51 obtained with our naïve approach). We can see that it achieved both a *higher*

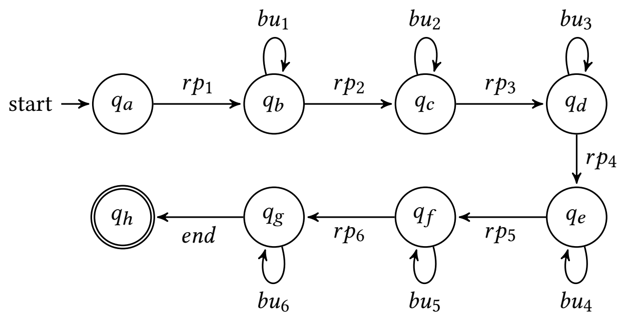

Figure 5: Finite State Machine representing the player's simulated behaviour (rp_x is the action of acquiring the game points for week x, bu_x is the action of buying an upgrade using the available game points during week x).

	ArmL	ArmR	Chest	Head	Legs
First week	1.01	0.98	0.97	1.02	1.01
Second week	1.14	1.11	1.12	1.13	1.14
Third week	1.21	1.13	1.12	1.18	1.19
Fourth week	1.16	1.04	1.05	1.16	1.12
Fifth week	1.16	1.11	1.15	1.13	1.11
Sixth week	1.16	1.22	1.30	1.17	1.04

Table 5: Mean pieces bought, divided by type and week, in the best solution (cost calibration) found.

mean number of pieces bought, meaning that more players were able to complete their robot, and a *lower* variance, meaning that the difference in robot completion between players was lower.

In particular, we conducted a Levene's test [27] for evaluating if the two samples have equal variances. We choose it as the skewness of the distributions prevented us to use a statistical test that assumes a normal distribution. The test obtained a f-ratio value of 7.20 and p-value of 0.0083. We can thus reject the null hypothesis at $p < .01$, and conclude that the novel approach was successful in reducing the variance of the game experience.

We report the distribution of the pieces bought at the end of the game by the simulated players in Figure 6. The percentage of classes that didn't buy a minimum of 10 pieces was 5.9 % (compared to the 26.6% of the previous calibration). The percentage of classes that completed the robot before the end of the project was 0% (identical to the 0% of the previous calibration). Finally, the percentage of classes that completed the robot, so obtaining a full game experience, was 31.96% (compared to the 40% of the previous calibration).

We can see that our novel approach was able to better achieve the game design objectives of ensuring a more even game experience. We intend to use the solution to future iteration of *WEEE R Robots*.

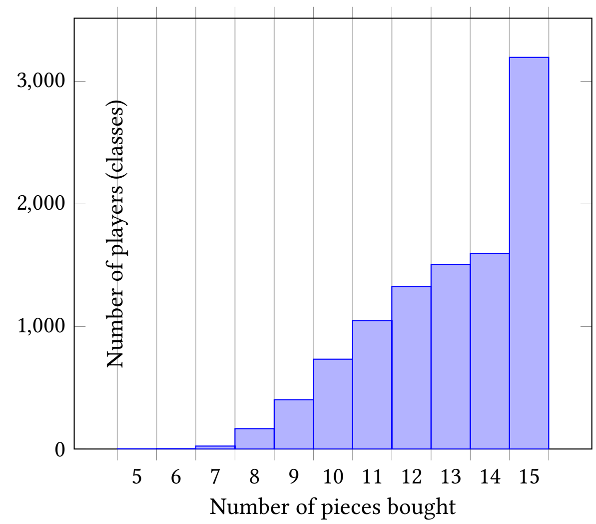

Figure 6: Distribution of the final number of upgrades bought in the best solution (cost calibration) found.

6 Conclusion and future works

In this paper we presented a novel approach for the calibration of game dynamics in multi–player games with a purpose. In this kind of games, players' retention is particularly relevant because maintaining the user engaged for longer may result in a more significant behavior change. To do so, not only the game needs to be carefully designed, but it also needs to adapt according to the player's situation and performance. In the context of *WEEE R Robots*, an educational game targeting primary school children and aiming at raising their awareness towards the reduction and correct disposal of EEE, game adaptivity is a particularly challenging problem. In *WEEE R Robots*, we aimed to promote an even player experience across classes, without overly rewarding high-performing players or penalizing low-performance ones, because low performance did not necessarily mean low children's commitment.

The approach we propose is based on the players' performances recorded during a past execution of *WEEE R Robots*. Here, we simulated the players game experience, and aimed to maximize its uniformity through a machine learning approach. We show that using the calibrated game elements we were able to obtain a more uniform game experience.

A limit of the proposed approach is that it assumes that the players' actions are static and unchanged for different calibration methods of game dynamics. This assumption could not hold in real-world cases, as players may change their behaviour according to changed game contexts. We plan to further study this limitation in a future work.

This suggested approach can be re-applied to other games in specific contexts, such as multi–player non–competitive or collaborative games, where the game designer may desire to minimize this variability, reducing the difference in game progression between low-performing and high-performing players. In order to do this it would be necessary to adapt the finite state machine representing the player's simulated behaviour, incorporating the specific different game mechanics or objectives. We plan to show such additional application of the approach as future work.

The proposed solution, along with further implementations such as the possibility to donate credits to the other classes, will be used and evaluated in the future trials of *WEEE R Robots* in the primary schools of Trento and might be of inspiration for other researchers and practitioners in the field dealing with game dynamics calibration in multi–player games.

Acknowledgments

This work is supported by the INNOWEEE project funded through the EU EIT Climate KIC under Project 190145.

References

[1] Shlomo Berkovsky, Mac Coombe, Jill Freyne, Dipak Bhandari, and Nilufar Baghaei. 2010. Physical Activity Motivating Games: Virtual Rewards for Real Activity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 243–252. https://doi.org/10.1145/1753326.1753362 event-place: Atlanta, Georgia, USA.

[2] Pavol Bielik, Michal Tomlein, Peter Krátky, Štefan Mitrík, Michal Barla, and Mária Bieliková. 2012. Move2Play: an innovative approach to encouraging people to be more physically active. In *Proceedings of the 2nd ACM SIGHIT international health informatics symposium*. ACM, 61–70.

[3] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[4] Kevin Browne, Christopher Anand, and Elizabeth Gosse. 2014. Gamification and serious game approaches for adult literacy tablet software. *Entertainment Computing* 5 (2014).

[5] Jenova Chen. 2007. Flow in games (and everything else). *Commun. ACM* 50, 4 (2007), 31–34.

[6] European Commission. 2018. WEEE Compliance Promotion Exercise. Final Report. https://publications.europa.eu/en/publication-detail/-/publication/09c7215a-49c5-11e8-be1d-01aa75ed71a1/language-en.

[7] Adam Crosier and Alison Handford. 2012. Customer journey mapping as an advocacy tool for disabled people: a case study. *Social Marketing Quarterly* 18, 1 (2012), 67–76.

[8] Mihaly Csikszentmihalyi. 1997. *Finding flow: The psychology of engagement with everyday life.* Basic Books.

[9] Mihalyi Csikszentmihalyi. 2014. *Flow and the foundations of positive psychology.* Springer.

[10] Simon Demediuk, Marco Tamassia, Xiaodong Li, and William L. Raffe. 2019. Challenging AI: Evaluating the Effect of MCTS-Driven Dynamic Difficulty Adjustment on Player Enjoyment.. In *ACSW*. 43–1.

[11] Simon Demediuk, Marco Tamassia, William L. Raffe, Fabio Zambetta, Xiaodong Li, and Florian Mueller. 2017. Monte carlo tree search based algorithms for dynamic difficulty adjustment. In *2017 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 53–59.

[12] Sebastian Deterding, Miguel Sicart, Lennart Nacke, Kenton O'Hara, and Dan Dixon. 2011. Gamification. using game-design elements in non-gaming contexts. In *CHI'11 extended abstracts on human factors in computing systems*. ACM, 2425–2428.

[13] Michela Ferron, Enrica Loria, Annapaola Marconi, and Paolo Massa. 2019. Play&Go, an Urban Game Promoting Behaviour Change for Sustainable Mobility. *Interaction Design and Architecture(s)* (2019).

[14] Kathrin Maria Gerling, Matthew Miller, Regan L. Mandryk, Max Valentin Birk, and Jan David Smeddinck. 2014. Effects of Balancing for Physical Abilities on Player Performance, Experience and Self-esteem in Exergames *(CHI '14)*. ACM, New York, NY, USA, 2201–2210. https://doi.org/10.1145/2556288.2556963

[15] Juho Hamari, Jonna Koivisto, and Tuomas Pakkanen. 2014. Do persuasive technologies persuade?-a review of empirical studies. In *International Conference on Persuasive Technology*. 118–136.

[16] Danial Hooshyar, Moslem Yousefi, and Heuiseok Lim. 2018. A Procedural Content Generation-Based Framework for Educational Games: Toward a Tailored Data-Driven Game for Developing Early English Reading Skills. *Journal of Educational Computing Research* 56, 2 (April 2018), 293–310. https://doi.org/10.1177/0735633117706909

[17] S. Howell and T. Veale. 2006. Dynamic difficulty adjustment in game-based intelligent tutoring systems. In *9th International Conference on Computer Games Artificial Intelligence and Mobile Systems*.

[18] Robin Hunicke. 2005. The case for dynamic difficulty adjustment in games. In *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology*. 429–433.

[19] Ruud S. Jacobs. 2016. Play to Win Over: Audiences and Effects of Persuasive Games. https://research.utwente.nl/en/publications/play-to-win-over-audiences-and-effects-of-persuasive-games

[20] Tomislav Jagušt, Ivica Botički, and Hyo-Jeong So. 2018. Examining competitive, collaborative and adaptive gamification in young learners' math learning. *Computers & Education* 125 (Oct. 2018), 444–457. https://doi.org/10.1016/j.compedu.2018.06.022

[21] Maurits Kaptein, Boris De Ruyter, Panos Markopoulos, and Emile Aarts. 2012. Adaptive Persuasive Systems: A Study of Tailored Persuasive Text Messages to Reduce Snacking. *ACM Trans. Interact. Intell. Syst.* 2, 2 (June 2012), 10:1–10:25. https://doi.org/10.1145/2209310.2209313

[22] Michael D. Kickmeier-Rust, Eva-C. Hillemann, and Dietrich Albert. 2014. Gamification and Smart Feedback: Experiences with a Primary School Level Math App. *International Journal of Game-Based Learning (IJGBL)* 4, 3 (July 2014), 35–46. https://doi.org/10.4018/ijgbl.2014070104

[23] Jonna Koivisto and Juho Hamari. 2019. The rise of motivational information systems: A review of gamification research. *International Journal of Information Management* 45 (April 2019), 191–210. https://doi.org/10.1016/j.ijinfomgt.2018.10.013

[24] Joey J Lee, Pinar Ceyhan, William Jordan-Cooley, and Woonhee Sung. 2013. GREENIFY: A real-world action game for climate change education. *Simulation & Gaming* 44, 2-3 (2013), 349–365.

[25] Nikoletta Zampeta Legaki, Nannan Xi, Juho Hamari, and Vassilios Assimakopoulos. 2019. Gamification of The Future: An Experiment on Gamifying Education of Forecasting. https://doi.org/10.24251/HICSS.2019.219

[26] Katherine N Lemon and Peter C Verhoef. 2016. Understanding customer experience throughout the customer journey. *Journal of marketing* 80, 6 (2016), 69–96.

[27] H. Levene. 1960. *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling.* Stanford University Press. 278–292 pages.

[28] Antonios Liapis, Georgios N. Yannakakis, and Julian Togelius. 2015. Constrained Novelty Search: A Study on Game Content Generation. *Evol. Comput.* 23, 1 (2015), 101–129.

[29] Annapaola Marconi, Gianluca Schiavo, Massimo Zancanaro, Giuseppe Valetto, and Marco Pistore. 2018. Exploring the world through small green steps: improving sustainable school transportation with a game-based learning interface. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*. ACM, 24.

[30] Baptiste Monterrat, Élise Lavoué, and Sébastien George. 2017. Adaptation of Gaming Features for Motivating Learners. *Simulation & Gaming* 48, 5 (Oct. 2017), 625–656. https://doi.org/10.1177/1046878117712632

[31] Adina Negrușa, Valentin Toader, Aurelian Sofică, Mihaela Tutunea, and Rozalia Rus. 2015. Exploring gamification techniques and applications for sustainable tourism. *Sustainability* 7, 8 (2015), 11160–11189.

[32] Rita Orji, Regan L. Mandryk, and Julita Vassileva. 2017. Improving the Efficacy of Games for Change Using Personalization Models. *ACM Trans. Comput.-Hum. Interact.* 24, 5 (Oct. 2017), 32:1–32:22. https://doi.org/10.1145/3119929

[33] Rita Orji, Lennart E. Nacke, and Chrysanne Di Marco. 2017. Towards Personality-driven Persuasive Health Games and Gamified Systems. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 1015–1027. https://doi.org/10.1145/3025453.3025577 event-place: Denver, Colorado, USA.

[34] Brian Orland, Nilam Ram, Dean Lang, Kevin Houser, Nate Kling, and Michael Coccia. 2014. Saving energy in an office environment: A serious game intervention. *Energy and Buildings* 74 (2014), 43–52.

[35] David R. Michael and Sandra L. Chen. 2006. Serious Games: Games That Educate, Train, and Inform. (2006).

[36] U. Ritterfeld, M. Cody, and P.A. Vorderer (Eds.). 2009. *Serious games: Mechanisms and effects.* Routledge.

[37] Katie Seaborn and Deborah I Fels. 2015. Gamification in theory and action: A survey. *International Journal of human-computer studies* 74 (2015), 14–31.

[38] Miyuki Shiraishi, Yasuyuki Washio, Chihiro Takayama, Vili Lehdonvirta, Hiroaki Kimura, and Tatsuo Nakajima. 2009. Using individual, social and economic persuasion techniques to reduce CO 2 emissions in a family setting. In *Proceedings*

of the 4th International Conference on Persuasive Technology. ACM, 13.

[39] A. Streicher and J.D. Smeddinck (Eds.). 2016. *Personalized and Adaptive Serious Games*. Entertainment Computing and Serious Games. Lecture Notes in Computer Science., Vol. 9970.

[40] HALLIFAX Stuart, Audrey Serna, Jean-Charles Marty, and Elise Lavoué. 2019. Adaptive gamification in education: A literature review of current trends and developments. In *European Conference on Technology Enhanced Learning (EC^{TEL})*. Delft, Netherlands. https://hal.archives-ouvertes.fr/hal-02185634

[41] Hallifax Stuart, Audrey Serna, Jean-Charles Marty, Guillaume Lavoué, and Elise Lavoué. 2019. Factors to Consider for Tailored Gamification. In *CHI Play*. Barcelona, Spain. https://hal.archives-ouvertes.fr/hal-02185647

[42] Julian Togelius, Georgios N Yannakakis, Kenneth O Stanley, and Cameron Browne. 2011. Search-based procedural content generation: A taxonomy and survey. *IEEE Transactions on Computational Intelligence and AI in Games* 3, 3 (2011), 172–186.

[43] Julita Vassileva. 2012. Motivating participation in social computing applications: a user modeling perspective. *User Modeling and User-Adapted Interaction* 22, 1 (April 2012), 177–201. https://doi.org/10.1007/s11257-011-9109-5

[44] Paul Weiser, Dominik Bucher, Francesca Cellina, and Vanessa De Luca. 2015. A Taxonomy of Motivational Affordances for Meaningful Gamified and Persuasive Technologies. In *Proceedings of the 3rd International Conference on ICT for Sustainability (ICT4S) (Advances in Computer Science Research)*, Vol. 22. 271–280.

[45] G. N. Yannakakis and J. Togelius. 2011. Experience-Driven Procedural Content Generation. *IEEE Transactions on Affective Computing* 2, 3 (July 2011), 147–161. https://doi.org/10.1109/T-AFFC.2011.6