Investigating the Educational Effectiveness of Multiplayer Online Games for Children

Franca Garzotto
HOC-Hypermedia Open Center, Politecnico di Milano
Via Ponzio 34/5, 20133 Milano – Italy
garzotto@elet.polimi.it

ABSTRACT
Multiplayer games are becoming an important part of Internet use, and have been the subject of many theoretical and empirical studies. Still, relatively few researches investigate multiplayer Internet games that are designed for young children and for educational purposes. This paper focuses on the educational effectiveness of this class of systems, and introduces some heuristics for its evaluation. We also report an empirical study that involved eighty-five elementary school children and measured the educational effectiveness of an online multiplayer game. We used both learning benefits analysis and empirical assessment of our heuristics, comparing the findings of the two evaluation methods.

Author Keywords
Multiplayer game, on-line edutainment, social interaction, children, collaborative learning, heuristics, inspection, empirical testing.

ACM Classification Keywords
K.3.1 [Computers and Education]: Computers Use in Education – Collaborative Learning; H.5.3 [Information Interfaces and Presentation] Group and Organization Interfaces - Web-based interaction, Evaluation

INTRODUCTION
The educational potential of multiplayer interactive games is acknowledged by many empirical studies in HCI [3][6][7][10][11][19][21] and has its theoretical foundation in a number of classical and modern theories in cognitive science and pedagogy [17][25].

Gaming provides an emotionally driven experience where players may learn a variety of skills, e.g., to take information from many sources and make decisions quickly, to apply their knowledge for creative problem solving, to develop strategies for overcoming obstacles, to optimize performance within constraints.

Since learning has a social dimension and has its root in social interaction [20][23][25], the educational benefits of games are potentially even stronger in situations of social gaming, involving multiple players.

Still, it is probably self-evident that not all multiplayer interactive games are equally educationally effective [8], nor are equally appropriate for adoption in an ordinary educational setting.

To assess the “educational effectiveness” of an interactive game we can measure its learning benefits - knowledge, skills, or attitudes acquired by learners by effect of the game experience.

Unfortunately, this evaluation is typically quite complex and time consuming, and per se does not highlight the reasons, i.e., the design characteristics, that make a game experience conducive (or not conducive) to learning.

This paper attempts to build a deeper understanding of the general design properties that contribute to the educational effectiveness of a multiplayer interactive game, and may foster its adoption in an educational context. Although most of our general considerations may apply to social games in any technological environments and for any target, we focus here on Internet-based multiplayer games for kids. This is the field where we have a larger expertise; furthermore, differently from other game technologies, Internet is widely adopted in schools today and represents, at least in the short-medium term, the most feasible channel to exploit social games as teaching and learning tools in conventional educational contexts.

In the first part of the paper we propose a set of educational effectiveness heuristics; they can be used as design guidelines for creating educationally effective multiplayer games, or as evaluation criteria, e.g., to complement learning benefits analysis or to gain insights of design problems that may affect learning. The second part of the paper discusses an empirical study that involved eighty-five elementary school children and eight teachers. We evaluated the educational effectiveness of a multiplayer Internet game using both learning benefits analysis and the proposed heuristics. The two evaluations provided consistent findings, both ratting the game as educationally effective – a result that provides a preliminary validation of our heuristics.

WHICH FACTORS MAKE A MULTIPLAYER INTERACTIVE GAME EDUCATIONALLY EFFECTIVE?
Drawing upon a literature review and our team’s experience in creating edutainment experiences for children [5][9], we have identified four major interdependent factors
that contribute to the educational effectiveness of a multiplayer e-gaming experience: usability, content, enjoyment, and social interaction. We omit to discuss usability, which is a fundamental but obvious pre-requisite and a well known concept. Ease of use of a game’s controls and interface is a pervasive factor that impacts on all the other aspects of the user experience: if users must struggle or cannot adequately translate their intentions into in-game behaviors they will become frustrated, will have less fun, and will loose motivation for engaging with game content and with other users.

In the rest of the section, we discuss content, enjoyment, and social interaction, decomposing them into a set of sub-factors (or heuristics) that are easier to measure and to analyse.

Content
We use the term content in a broad sense, to denote both the information that players are exposed to during the game and the tasks that are required to complete the game. As observed in [8], the simple presence of “educational” content in a game does not guarantee its power for learning and its potential for acceptance and adoption in a school context. A number of characteristics may affect the educational effectiveness of game content, as discussed below.

Target/Goal Appropriateness
The content should be tailored to the developmental and instructional level of game users and to the educational goals of the context of use.

Integration
Educational content should be an integral component of the game: the act of play should draw directly on the knowledge and skills that the game is designed to foster in its users, and should promote reflection about or application of such knowledge and skills.

Scaffolding.
The feedback provided in response to incorrect actions should offer an opportunity for learning, exposing players to educational content and encouraging them to continue the game.

Extensibility
Learning is facilitated by the effective combination of different kinds of (digital and physical) resources, and different modes of developing knowledge and skills [20][25]. This paradigm, known as blended learning, can be translated into game content characteristics: concepts, topics, and tasks introduced by the game should promote off-line reflections and be appropriate for further off-line exploration and application; the game experience should be smoothly extensible to ancillary educational activities in the classroom or at home (e.g., reading, writing, drawing, story-telling, group discussions, …); game content should be “consistent” with the curriculum and the goals of the education context where the game is employed.

Media matching
Not all media are equally appropriate for all types of contents. A topic can be communicated more effectively using animation, another one using sound, another one using an interactive map; a task can be performed better in a 2D environment than a 3D space. Contents should be rendered using the most appropriate “media”.

Enjoyment
Enjoyment is central to any gaming activity: It represents a key motivation for engaging in a game, and is a crucial driver to involve students emotionally and to create an intrinsically rewarding learning experience. A wide literature exists on enjoyment in general, and on enjoyment in gaming in particular [12][17][24]. Our main source of inspiration is represented by Csikszentmihalyi’s Flow Theory [2]. Founded on a massive set of ethnographical studies, this theory identifies the elements of enjoyment that are universal, irrespective of age, social class, culture, or gender. We have tailored this and other approaches (see section “Related works”) to an educational perspective, defining some interdependent components of enjoyment that we consider particularly conductive to learning in multiplayer online games for kids.

Clear goals
The game should provide the players with clear goals at appropriate times, to help them concentrate on the game activities, develop the proper strategies, and finalize their actions.

Concentration.
The game should keep the player’s concentration through an appropriate work-load and proper stimuli, so that the person’s attention is completely absorbed by the activity, and all relevant skills are employed in it.

Challenge.
For many players, testing their abilities and overcoming obstacles is why they play [13][17]. Challenge not only creates emotions and rewards. From an educational perspective, it helps to focus attention, inspires creativity, and fosters the development and application of skills or strategies. To be “sufficiently” challenging, game tasks should match the player skill level without being discouragingly hard or boringly easy; they may progress thorough a series of challenges requiring a progressive level of skills, unlocking new challenges and opportunities along the way.

Feedback.
Players must receive feedback not only on interaction effects (which is a general usability principle) but also on success or progress towards completing game tasks. Enjoyment increases when feedbacks are only just explicative, but also create wonder or surprise.

Immersion.
Players should feel totally absorbed in the game, feeling effortless involved in the gaming activities, becoming less self-aware, experimenting an altered sense of time, and losing awareness of the surroundings.

Social Interaction
The social basis for learning, particularly in early years and primary education, has been acknowledged since the seminal research of the Russian psychologist Vygotsky
In the context of gaming, social interaction represents both a vehicle to learn collaboration attitudes and skills, and a powerful motivator to engage in educational content. Various forms of social interaction may promote the educational effectiveness of a game: connection, cooperation, and competition.

Connection
The game should create the feeling of “being there with many” (or “co-presence” [14]), of being “a member of a group” (community) and promote connection with other people inside or outside a game: these features enhance the sense of social identity and enforces the feeling of self-recognition without which learning attitudes, in kids, may be weakened [25]. The game should support means to discuss strategies and solutions with others, which is a way “to make thinking visible” and to create a more tangible context for what has been learned, building the conditions for situated learning [12].

Cooperation.
For educational effectiveness, cooperation should take two forms in a multiplayer on-line game:

- **Remote cooperation** with distant peers (implicit in the concept of “multiplayer online” game). The game should have clear shared goals that pinpoint the need for more than one person to be involved and for complementary skills and knowledge; the design should support the acknowledgment of the mutual benefits of working together (e.g., through appropriate feedbacks). Cooperation should not affect pace (the rate at which players experience new challenge). The need of coordination with others should not force a suspension of individual activities: kids should not be forced to wait inactively “until others finish something” but should be enabled to continue to experience novel game details.

- **Co-located cooperation** among shoulder-to-shoulder mates. In the ordinary school environment, children usually learn by interacting with each other while in one place or in a face-to-face situation. Computers are seldom used individually, but there are typically 2-3 kids working together with a single computer. Co-located cooperation is an important issue, and the game should be designed to foster this aspect.

Competition
Most people, kids in particular, gain satisfaction from competing against and beating peers. Although competition is in antithesis of cooperation, the need or desire to compete with others is a very common impetus that motivates individuals to cooperate with each other in order to form a stronger competitive force. Competition adds emotions that make the gaming experience more exciting and fun [13][24], and is a powerful motivator for being engaged in educational content, both during the game, and before the game, e.g., in preparation for the “race”.

EVALUATING EDUCATIONAL EFFECTIVENESS: A CASE STUDY
The goal of our study was to test the efficacy of our heuristics for the evaluation of educational effectiveness. We considered an online multilayer game developed in the context of a larger project [9] that concerns internet based edutainment for elementary school children. The application is implemented using Flash Communication Server and XML technology.

Our study evaluated our educational effectiveness heuristics (by means of a variety of techniques – observation, chat log, questionnaire, interview) and also empirically measured, by means of learning tests, the learning benefits achieved by 85 kids using the game. We finally compared the results of the two approaches. For learning benefits, we focused on two levels of Bloom’s taxonomy of the learning domain [1]: the cognitive level (which involves knowledge and intellectual skills) and the affective level (which involves the manner in which learners deal with things emotionally - feelings, values, appreciation, enthusiasms, motivations, and attitudes). For learning benefits at the cognitive level, we focused on the cognitive skills of recall, recognition, and understanding of specific facts or concepts. For the educational benefits at the attitude level, we considered: motivation to learning new subjects, interest to different cultures, willingness and satisfaction in cooperation.

The Game
“Pirates Treasure Hunt” (“Pirates” for short) is a discovery game that has the educational goal of stimulating interest and attention on “other cultures and ways of life” - a subject of primary school curricula in Italy. Pirates exposes kids to contents related to non-European countries; it requires players to apply (previously studied) knowledge about these cultures as a pre-requisite to be able to complete the game tasks. At a deeper level, Pirates fosters concept understanding and meaning-making, as well as recognition and recall skills. In the game, players act like pirates, who explore a virtual archipelago and discover hidden “treasures” with the help of remote players. In the virtual world, a user is represented as a small ship and navigates either in a continuous way (using cursors) or “by jumps” (selecting an area on a map).

The archipelago is populated with interactive objects - animals or environmental elements. Some of them (e.g., rocks and sharks) are “dangerous”: when casually encountered or explicitly activated, they cause a penalty and a brief (10 seconds) suspension of the player activity, during which any interaction (but chatting, discussed later) is inhibited. Other objects hide “treasures” that appear when objects are clicked. A treasure is a quiz card that shows the image of a typical cultural aspect of a country (e.g., a typical food of Morocco or a typical piece of furniture of Japan), and a yes/no question about the image meaning, such as: “Is this Japanese food?” (see figure 1). The game supports eight simultaneous remote users who are organized in two competing teams, each one representing a country (Morocco or Japan). Each player is assigned to a mission: discover and collect all cards (“treasures”) that concern a specific aspect of the culture of his country team (food, dressing, tradition, or furniture).
The set of treasure cards hidden in the archipelago is the same for all players, but the card quiz appearing to a player is dynamically customized to his mission. For example, a treasure card may contain an image of sushi food but the question will be “Is this Japanese food?” for the player whose mission is to collect Japanese food, but will be different for another player. If the answer is wrong, the player gets a penalty, and a 10 seconds suspension. If the answer is correct and the image corresponds to the player mission, the card is collected and contributes to a team’s progress towards victory.

The game ends when all treasures (i.e., cards) for a particular country are discovered, and victory is assigned to the team whose members all completed their individual mission.

This collective goal is the key motivator for collaboration and communication. A player should feel the other’s missions as important as his mission, and be motivated to continue searching for team treasures and to offer mutual assistance even after his individual mission is completed.

The communication and collaboration tool is the chat. A player can send messages to invoke or provide help, to alert the other members about pitfalls or dangerous situations in a given area, or to immediately inform his team mates when he discovers a treasure that is relevant for his team’s overall goal.

The game tasks require kids to have previously acquired some factual knowledge about the game subjects. We have created ancillary printed material that can be studied in the classroom or at home and comprises information sheets about the various topics, with images and texts describing the subjects of the treasures hidden in the game. We have also created a 3D multimedia learning space that kids can use individually.

Figure 1: Pirates’ Interface

Figure 2a: The 3D Learning Space: a Japanese House

This system provides cartoon-like 3D worlds that reproduce “typical” environments of Morocco or Japan (see figure 2a), where the “cultural elements” of each country (examples of food, dressing, traditions, furniture that represent the game treasures) are shown “in context”.

When an element is selected, the corresponding information sheet appears (see figure 2b). The images for the objects in the treasure cards of the game, in the 3D learning environment, and in the printed sheets, are all the same.
Participants
The study involved eighty-five kids from a public elementary school in Milan. The participants comprised 44 girls and 41 boys; 65 kids were 7-8 years old (from three second grade classes), and 20 kids aged 9-10 (from a fourth grade class).

All kids had a sufficient level of skills in the use of mouse and keyboard, and, in spite of age differences, a pretty homogenous exposure to interactive experiences (for various reasons, second grade classes had participated in a larger number interactive multimedia projects than their older mates).

Eight teachers (two per class) were involved in the study; they offered assistance during the game sessions and were interviewed afterwards.

Environment and Collaboration Setting
We conducted 4 test sessions (each one of an approximate duration of 115 minutes). Each one involved an entire class (20-23 kids) and was held in the two computer laboratories of the school.

In each session, we organized kids in two teams (one for Japan and one for Morocco) and further subdivided a team in four groups.

Each group, composed of 2 or 3 kids (depending on the class size), was assigned to a computer (see figure 3) and represented a single player. Teams and groups were homogeneous in terms of genre distribution and intellectual ability.

To simulate the effect of “remote” collaboration among groups/players, we placed two groups per team in each lab; within the same room, computers were distant enough so that kids were not distracted by the activities of others collocated groups.

Kids' Activities during a Test Session
For the first 10-15 minutes, kids refreshed their knowledge about game topics and used the 3D learning environment then they were trained about game rules and spent approximately 10 minutes to practice with interaction commands on a demo-version of Pirates.

Finally, they played the real game twice, changing country and mission (but not team and group).

Data Collection Techniques
We used various techniques for qualitative and quantitative data collection:

- Behavior Observations and Chat Logs. An observer for each group sat side-by-side kids and took notes on a structured evaluation form. Play sessions were also video recorded (using two moving cameras).

- Questionnaires; a set of closed and open-ended questions were submitted to kids immediately after a test session (figure 4). They asked kids what they liked and disliked during the game, how easy and pleasant the group collaboration was, whether or not they would like to repeat the experience, and similar. A “Smilemyometer” ☺☺☺☺ was used as metric for the kids [18]

- Learning tests. To measure learning benefits at the knowledge level we submitted a learning test before and after the gaming experience. Knowledge and intellectual skills (recognition, recall, concept understanding) were operationalized in terms of kids’ capability to assign meaning to images. Each test included ten multiple answer closed questions (“quizzes”). A quiz showed an image (e.g., sushi food) randomly selected in the set of the game.
treasure cards and proposed four alternative solutions for its country and “type” (food, tradition, furniture, or dressing).

- Interviews to the teachers. They were held some weeks after the test sessions, and mainly attempted to get insights about appropriateness of contents and learning benefits at the attitude level.

**FINDINGS**

We used quantitative techniques to analyse pre-post test scores and closed questions answers in kids’ questionnaires. Qualitative data analysis methods were applied to verbal and visual data (interviews transcripts, observers’ notes, video recordings, chat messages, kids’ answers to open ended questions).

Our findings indicate that Pirates does produce some learning benefits. For the knowledge domain, the main metrical was the number of correct answers in the pre and post tests. A paired t-test shows that the average number of correct answers before and after the gaming experience is respectively $M_{before} = 5.14$ and $M_{after} = 7.16$, showing a statistically significant increment of $2.1 = 29\%$ correct answers ($t_{(10.25)} = 0.01$).

Concerning *integration* (final score = 3), we considered the number of situations in which kids, to answer treasure quizzes, tried to merely guess, rather than to use their knowledge about the subject matter. Analysis of observers’ notes reveals that guess occurred in 12\% of the cases. In all other cases kids tried to reason and to discuss (see also section “Group Cooperation”)

**Enjoyment**

The overall enjoyment was very high, as figure 5 highlights. In addition, the number of free comments in kid’s questionnaire denoting the degree of enthusiasm and appreciation resulted highly enthusiastic in 82\% of the cases (“I would like to do it thousand times!!!”, “I wish I had it at home thus I could play every day!”); 15\% were “rather” enthusiastic, 2\% neutral, and 1\% negative. Concerning *clearness of goals* (final score = 4), observers’ notes show that for 96\% of the groups it was immediately clear what they had to achieve, and 4\% “understood after a while (3-5 minutes)”.

For concentration (final score = 3), observers’ notes and video recording reported that the game almost constantly grabbed kid’s attention. Some distractions were noticed when kids encountered dangerous objects (because of the suspension); some players learned that they could chat in the meanwhile, and used suspension time for concentration.

Concerning *well-being* (final score = 4), teachers used the game cultural stimuli in classroom and homework (discussion, storytelling, drawing). They found the 3D learning space particularly motivating and inspiring for the purpose of these activities, suggesting that an ancillary learning space about a game subject can be very useful for enforcing blended learning.

Figure 5 - Kids’ Questionnaire: key results (Left diagram: Kids’ preferences; Right diagram: satisfaction)

For learning benefits at the affective level, our data highlighted that kids appreciated “learning new things about other cultures”: this aspect was the second highest scored among five multiple choices for “What do you like more?” (see Figure 5). Teachers’ interviews confirmed this result, reporting significant interest and motivation in the classroom discussions and activities on game subjects that were carried on after the Pirates laboratory sessions.

For the evaluation of our heuristics, we classified the overall information collected during test sessions and teachers’ interviews according to their pertinence to the different heuristics; we then coded each contribution as denoting a certain degree of heuristics fulfillment (4 = very high, 3 = high, 2 = medium, or 1 = low), finally calculating the means to derive the final scores (appearing in brackets in the rest of this section). The evaluation of our heuristics confirmed the educational effectiveness highlighted by learning benefits analysis, as discussed below.

**Content**

Teachers judged game tasks and topics as *target appropriate* (final score = 3) and at the proper level to address multicultural issues.

Concerning *extendibility* (final score = 4), teachers used the game cultural stimuli in classroom and homework (discussion, storytelling, drawing). They found the 3D learning space particularly motivating and inspiring for the purpose of these activities, suggesting that an ancillary learning space about a game subject can be very useful for enforcing blended learning.
boredom. The questionnaire results highlighted that “discovery” was the most appreciated aspect of the game (receiving the highest score by 68% of kids among – see figure 5). Only one kid explicitly reported that he “wanted to have more dangerous objects to discover”. All players were eager to continue the session also after completing their mission.

For immersion (final score = 4), kids were not distracted by the surroundings: they never moved from their place, nor talked about subjects other than the game. They did not realize when the bell rang and it was break or lunch time. A kid affected by the Tourrette Syndrome (a serious neurological disorder characterized by the presence of multiple motor and phonic tics) was so immersed in the game that none of his syndrome symptoms manifested (we discover his disease only later, from teachers’ interviews).

Social interaction
For group cooperation (final score = 2), among co-located kids, kids cooperated well to find the right solutions to treasure quizzes (discussing and taking decisions together). Still, while 40% of groups worked in harmony for the whole test session, in 60% of the groups some conflicts emerged: kids did not want to alternate in the control of game interaction.

Approximately at the middle of a game session, these groups found an explicit or implicit agreement about members’ roles (e.g., one member chatting and writing, another one exploring the world, the third one sitting in the background for a mainly intellectual role).

The questionnaire results are consistent with field observations. Group cooperation was the less appreciated aspect of the game (being selected only by 18% of kids as an aspect they liked). Other answers reveal that group cooperation was perceived as “difficult” by 27% of kids, “rather difficult” by 39% of kids, and “easy” by 34%. In contrast, teachers reported that the overall experience was beneficial for improving cooperation attitudes: in following conventional laboratory sessions, kids manifested a more disciplined cooperation behavior than they were used to before the Pirates experience.

In average, kids scored Remote Cooperation (final score = 3) as the third most appreciated aspect, together with Communication (final score = 3) - see values of “Chatting” and “Playing with my team” in figure 5). In both cases, 25% of them picked these two choices as something they liked.

Remote cooperation is the only aspect in which the results significantly diverge between different ages. A higher appreciation of remote cooperation occurred in older, fourth grade kids: 65% of them selected it as something they liked, against 18% of younger second grade kids. Five of older kids also reported very positive spontaneous comments about remote cooperation in the questionnaire (e.g., “it was so nice to play with my friends without even seen them!”) while this aspect was never mentioned in younger kids’ written or spoken comments.

As we noticed before, all players showed keen interest and intense desire to continue playing also after they had completed their mission, and cooperation became stronger after that time. Chat logs and field observations highlight that the use of chat became more intense after a group completed its mission, and in 70% of the groups actually started only after that time!

Once again, cooperation was higher in older kids (who used the chat for this purpose 30% more than their younger mates). These differences may be due to a better capability of fast computer writing in older kids: 60% of them estimated “easy” the use of the chat, against 32% of younger kids. Chat messages concerning topics other than the game were very limited (5%), probably due to the pressure imposed by competition. An interesting aspect was the language that some groups developed to optimize the communication with team mates. While at the beginning they used full phrases (e.g., “warning! A grey shark is in area 3D” or “Search for Japanese food near the mountain”) they progressively minimized their language, dropping articles, verbs, adverbs, and adjectives. Thus the previous sentences become: “shark 3D”, or “food – mountain”.

Competition (final score = 4) was definitely an enormous driver for engagement and enjoyment. The happiness of victorious kids was enormous (they shouted and jumped for 5 minutes), but also the players of the defeated team did not loose their enthusiasm, immediately asking for playing again.

Competition fostered excitement: the progress bar was kept under constant control by the whole group, and a progress of the own team was outlined by enthusiastic comments.

Competition fostered cooperation: chatting and engagement increased when kids realized, on the progress bar, that the other team was getting close to victory. Creative strategies were found under the pressure of competition, like the language minimization mentioned above.

RELATED WORKS AND CONCLUSIONS
Several works have developed game heuristics [4][15][22] that address both general usability aspects and game specific aspects, such as enjoyment. Still, none of them looks at games from an educational perspective, nor focus on kids. A notable exception is [8], but this work considers content aspects only.

Existing approaches use game heuristics as design guidelines or criteria for expert review, specializing Nielsen’s general method of usability inspection. Still, educational effectiveness heuristics are difficult (if not impossible) to measure through expert review only, especially in the children domain. Their evaluation requires a variety of data of different nature, both quantitative and qualitative, which can only be collected through empirical methods, involving players (kids) and teachers.

Most of HCI empirical research on online multiplayer games does not address young kids, mainly focusing on adults [6][11] or adolescents [5] (and marginally
considering educational aspects). For pre-school or primary school kids, empirical results on social games and learning consider technological settings (tabletops, tangibles, robots, mobile devices, augmented reality systems, etc.) [7][16], which are usually unaffordable by a typical educational environment.

The main contribution of the work reported in this paper is to offer to game designers, evaluators, and teachers, some methodological and empirical results for kids’ game based learning in a technological domain – Internet - that is becoming more and more popular in our schools.

Our study also gives some practical insights on how to empirically measure educational effectiveness of online social games for children, paving the ground towards the definition of a more general evaluation framework for this class of systems.

Acknowledgments

The author is grateful to the children and teachers of classes 2a, 2b, 2c, 4d at the primary school “Nolli Arquati” in Milano who participated in our evaluation study. A special thank to the HOC team for the development of “Pirates”.

REFERENCES