

027607

mGBL

mobile Game Based Learning

Specific Targeted Research Project

Information Society Technologies

D7.3 Report on evaluation of 1st user trials and elaborated games [revised version]

Due date of deliverable: 31. October 2006.
Actual submission date: 8. November 2007.

Start date of project: 1 October 2005
Duration: 36 months

Version 1.1

Project co-funded by the European Community within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	X
PP	Restricted to other program participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Table of Contents

1	Introduction	4
1.1	General Deliverable Description	5
1.2	Revision history of this document	5
1.3	External peer-review (internal chapter, not part of the public deliverable)	6
1.3.1	<i>General description of the review process.....</i>	6
1.3.2	<i>Comments and Recommendations of the External Reviewers</i>	6
1.3.3	<i>Justification and actions/corrections taken regarding the external peer-review.....</i>	6
1.4	Executive Summary of the Deliverable	7
2	Evaluation of methodology applied	8
3	Global evaluation of student trial results	9
3.1	Evaluation of initial review results	10
3.1.1	<i>Computer experience</i>	11
3.1.2	<i>Mobile phone experience.....</i>	11
3.1.3	<i>Learning habits</i>	13
3.2	Evaluation of general first impressions	15
3.2.1	<i>Opinion about the games.....</i>	15
3.2.2	<i>Liking of the games.....</i>	16
3.2.3	<i>Disliking of the games</i>	16
3.2.4	<i>Feelings during the game.....</i>	17
3.3	Evaluation of students' ratings for Game 1	18
3.4	Evaluation of students' ratings for Game 2	19
3.5	Evaluation of students' ratings for Game 3	20
4	Global evaluation of teachers' trials results	22
4.1	Evaluation of general first impressions	22
4.2	Evaluation of teachers' ratings for Game 1	23
4.3	Evaluation of teachers' ratings for Game 2	25
4.4	Evaluation of teachers' ratings for Game 3	27

4.5 Pedagogical concepts of the games: a reasoned summarise of teachers' and guidance practitioners' comments after the first users' trials	28
4.5.1 <i>Context.....</i>	28
4.5.2 <i>Positive perception of the games.....</i>	29
4.5.3 <i>Negative perception of the games</i>	31
5 General analysis and cross comparisons.....	34
6 Suggestions for improvement of the games.....	38
7 Suggestions for next trials.....	40
7.1 Methodology	40
7.1.1 <i>Introducing User Trials</i>	40
7.1.2 <i>Different Types of Usability Tests</i>	40
7.2 Application to mGBL – Project.....	44
7.2.1 <i>Second User Trials</i>	44
7.3 Schedule for User Trials	44
7.3.1 <i>Planning for Testing</i>	44
7.3.2 <i>Preparing for Testing.....</i>	45
7.3.3 <i>Conducting the Test</i>	46
7.3.4 <i>Analysis of Results.....</i>	46
7.4 Reporting the Results.....	47
8 Conclusions.....	48

1 Introduction

The first user trials have been conducted online with the respondents from 5 partner countries (Austria, Great Britain, Italy, Slovenia and Croatia). Two groups of subjects undertook the trial; students (age 18-25) and teachers. Survey took place from 23rd to 31st of October 2006.

Results from 137 students and 13 professors were collected.

The original trial concept had altered due to the WP rearrangements as well as overlapping with exams and summer vacation. Therefore, the original time scheduled for the user trials was postponed and shortened.

The complete 1st user trial took place during one week, and the report on it followed the subsequent week.

According Annex I – “Description of Work” of mGBL Project, the Work Package 7 has the objective to evaluate the user trials, in particular methodology and feedbacks provided. The following report is the evaluation of methodology and results of the 1st user trials conducted under the coordination of Work Package 6. The main input for this document was the “D6.2 First user trial report” that gathers together the first user trial data.

There were three games tested: Game 1 “On the edge”, Game 2 “Ahead of the game” and Game 3 “Baroque Blog”.

This section provides general information about the deliverable.

They are:

- General Deliverable Description
- Revision history
- External peer-review (internal, not part of the public deliverable)
- Executive Summary of the Deliverable

1.1 General Deliverable Description

WP number:	7
WP name:	Evaluation and Validation
Deliverable number:	D7.3
Deliverable name:	Report on evaluation of 1st user trials and elaborated games
Responsible work package leader:	
name:	Walter Ukovich
address:	via Valerio 10, Trieste
email:	ukovich@units.it
phone:	-
mobile:	-
fax:	+39 040 558 3414
Involved project partners:	PFRI, FFRI, Aster, UM, AZM-LU

Table 1: General Deliverable Description

1.2 Revision history of this document

Date	Version	Description	Author
31/03/07	1.0	Final version	TRIESTE
08/11/07	1.1	Revised version	TRIESTE

1.3 External peer-review (internal chapter, not part of the public deliverable)

This section contains a description and an overview of the results of the external peer-review of the deliverable. This is an internal chapter (consortium, Project Officer and reviewers) and will be removed within the final public version of the deliverable.

1.3.1 General description of the review process

This deliverable was reviewed by prof. G. Vercelli.

Reviewer profile: Gianni Vercelli received his Laurea degree in electronic engineering in 1987 and his Ph.D. in computer science in 1992. He was with the University of Trieste, Italy, from 1996 to 1999, and he is currently an Assistant Professor in Computer Science and Multimedia Design at the Education Faculty of the University of Genoa. He is a member of the IEEE Computer Society and of the Italian Association for Artificial Intelligence. His scientific interests are focused on robotics and artificial intelligence, intelligent agents, and multimedia education. He has written more than 70 papers.

1.3.2 Comments and Recommendations of the External Reviewers

Comments by prof. G. Vercelli on this supplement:

Supplement is clear enough. Take care to separate analysis of results from evaluation of methodology. In some points the two views overlap.

1.3.3 Justification and actions/corrections taken regarding the external peer-review

Modifications to original version have been made. Similar remarks as from external reviewer were made by EU reviewers. Care was taken to separate evaluation of methodology from evaluation of results. Yet evaluation also of results is an important part of the tasks of WP7. Although there must not be overlap with work of WP6, one of the tasks of WP7 is to verify "Empirical

evidence of the learning effectiveness, usability and acceptability of the mGBL 1st test applications.”

1.4 Executive Summary of the Deliverable

This deliverable D7.3 gives evaluation of the 1st user trials and elaborated games. Work Package 7 has the objective to evaluate the user trials, in particular methodology and feedbacks provided. The following report is the evaluation of methodology and results of the 1st user trials conducted under the coordination of Work Package 6. The main input for this document was the “D6.2 First user trial report” that gathers together the first user trial data.

2 Evaluation of methodology applied

In the D6.2 “First user trial report” introduction it’s explained that candidates who will participate at user trials will be selected but not how the actual respondents were selected. Testing the prototype games was the major point of user trials but **it’s not clear how students and teachers tested the three games** (where, for how long, alone or in group, on-line or on mobile phone); it’s only explained how the on-line questionnaires were set. It was also crucial that all tests would be done under controlled conditions not as an online test.

At the end of D6.2, are clearly laid out questions that were addressed during the first user trials. **Questionnaires’ design is not justified enough:** it should be indicated why these questions were selected and, if relevant, the project participating organizations that consider them important. If evaluation questions that some might expect to see were eliminated, the reasons for eliminating them should be provided. In the questionnaire for students at the end of user trial were presented the same questions that were asked in the initial review survey: in D6.2 this fact is not motivated at all.

During these first user trials there has been only one structured group discussion with 27 students in Rijeka. Answers given by the students seem to be very synthetic and out of topic (e.g.: Q. Can games give value to learning process? A. Yes Simulators are good example. Q. What are the weaknesses of games? A. Insomnia, cannot stop playing.). **The usefulness of this group discussion is questionable.** It’s not clear how the group discussion was conducted, by whom, etc... During a structured group discussion, everyone should be invited in turn to offer two or three sentences about the issues. Every opinion should be noted down: we are interested in the range of views, and that we don't expect them to reach agreement.

3 Global evaluation of student trial results

The evaluation of students' trial results is based on the answers and comments given by 137 students (93 male, 44 female) from the following institutions:

Anglia Ruskin University (GB)

Karl Franzes University Graz (AUT)

Faculty of Maritime Studies in Rijeka (HR)

University of Bologna (IT)

University of Trieste (IT)

University of Ljubljana (SI)

University of Maribor (SI)

The questionnaire filled in at the beginning and at the end of the game testing according to D6.2 First User Trial Report was evaluated. Responding students (subjects) had the following majors: business administration, chemistry, computer science, economics, engineering (electronic and clinical), electronics and communications, geography, logistics, pedagogy, science technology, social sciences, sociology.

Questionnaire consisted of students' first impressions of the game; the subjects were asked for their opinion about the game, what did they like or dislike, were their prior expectations met, whether the stated objectives were met and their individually expected goals reached. Students were also asked to state their feelings during the play and to rate their answers to the questions referring to individual games. They chose among the following responses: excellent, good, sufficient, poor, and very poor. The responses "excellent, good and sufficient" are considered as different levels of positive comments and "poor" as well as "very poor" as negative.

3.1 Evaluation of initial review results

At the beginning of test, students received a questionnaire. The topics of the questionnaire are:

- Computer experience,
- Game experience,
- Mobile phone experience and
- Learning habits.

In brief, this survey contains several interesting findings: students want technology in learning more than teachers (60% vs. 40%), high school students use mobile phones more than university students (except phoning), mobile phones are rarely used in learning (before trial), the average computer usage is 5 hours/day, the average internet usage 3 hours/day, playing 1 hour once per week.

Nevertheless, some remarks should be done. **The results are presented grouped for countries and there are not the numbers of the effective persons who answer: this makes harder comparing results. It would be more useful to present the real amount of respondents, analyzed all together. Gender issues were not analysed: it would be interesting if results would be presented also grouping female and male answers.**

We should take care of research work that is done elsewhere and that is available. These research works can give a concrete validation of mGBL outcomes.

The European Commission regularly carries out public opinion surveys in order to keep abreast of the consumer side of the rapidly evolving telecommunications markets and, in particular, to assess how consumers benefit from the liberalization of the market.

The data below are extracted from a survey that was carried out between 17 November and 19 December 2006 and it follows on from the Eurobarometer study that was conducted between December 2005 and January 2006.

3.1.1 Computer experience

Single households by age with at least one computer (Eurobarometer)

EU27			EU15			NMS12		
-29	30-59	60+	-29	30-59	60+	-29	30-59	60+
72%	53%	15%	73%	57%	17%	70%	29%	6%

According to the mGBL survey young people who has at least one computer, are about 98%. Respondents to mGBL survey were university student while sample analyzed by Eurobarometer was not such limited. It could be inferred that young people who study at the university are significantly more likely to have a computer than others.

From Eurobarometer report, 42% of households within the EU27 have internet access at home. A slight increase can be observed within the EU25 (44%) since the winter 2006 study (+4 percentage points). The technology used to access the internet varies widely between the countries. At EU level we can broadly say that about two-thirds of the households that have internet access use broadband technology while the remaining third use narrowband connections. In the total sample of households, 28% have a broadband connection and 12% use a narrowband connection. The patterns observed for single households regarding the resident's age and whether they are equipped with telecommunication services or appliances also apply here: the younger the resident is, the more likely there is a broadband connection in the household.

3.1.2 Mobile phone experience

Single households by age with at least one mobile telephone access (Eurobarometer)

EU27			EU15			NMS12		
-29	30-59	60+	-29	30-59	60+	-29	30-59	60+
95%	83%	43%	95%	86%	46%	95%	70%	29%

These data agree with mGBL survey results: 96.27% of the respondents, who are university students 18-25 years old, have a mobile phone.

The Eurobarometer work didn't make an analysis of the mobile phone usage. Handset maker Nokia commissioned a survey aimed to find out just how much use people make of the ever-growing list of functions crammed into modern mobile phones. The study was published in June 2006.

Results of Nokia Nseries study reveal widespread consumer demand for digital convergence. Nearly half of respondents (44%) use their mobile device as their primary camera. Globally seventy two percent now don't use a separate alarm clock and 73% use their mobile as their main watch or clock. When it comes to surfing the web on the move, over a third (36%) of respondents is browsing on their mobile devices at least once a month.

Music-enabled mobile devices are set to be at the heart of future digital music consumption: 67% predict the mobile phone will replace their MP3 player.

Almost half (42%) of respondents globally want their printer, PC, stereo, TV and mobile device interconnected. And looking to the future, more than one in two (58%) of those questioned would like to be able to control all their household appliances via their mobile device.

With an ever increasing list of features, it's not surprising that people can't imagine life without a mobile; mobile are so indispensable that over one in five (21%) would be more upset about losing their mobile than wallet, credit cards and even their wedding ring.

The mGBL survey underlines that only 15% of the interviewed sample have never played games on mobile phone. Similar data involving young European people are not available, but other researches in mobile game field contain interesting findings that confirm the growth of mobile gaming habits.

According the NPD's (an American Market Research Firm providing news and information on consumer trends, sales and marketing information for a diverse range of industries) wireless research team, from a game content standpoint, 20 percent of mobile consumers played games on their handsets in 2006 in the United States. Just over 3 percent of mobile phone owners download a mobile

game in Q2 2006, and that percentage increased only slightly to just shy of 4 percent by Q4 2006.

The mobile game market is doing quite well. In Q4 2006 the industry took in nearly \$107 million in mobile-game-related revenue. This data-service revenue was topped only by short message service (SMS) revenues, which reached \$482 million for the quarter, and ring-tone downloads, which reached \$172 million.

In the United States, the most popular games are also those that are available in other non-mobile formats, from which it can be inferred that it remains difficult for mobile game developers and publishers to gain consumer traction with mobile games that are limited to the wireless arena alone.

Total global revenues from mobile games are forecast to increase from \$2.6 billion in the 2005 to \$11.2 billion by 2010, according to Mobile Game, a new strategic research report from Informa Telecoms and Media (a leading provider of business intelligence to global telecoms and media markets).

Downloads will account for around two-third of total global revenues through 2010, but online multiplayer traffic will start to generate significant income for mobile operators, as cellular operators launch more multiplayer games and introduce community features that will encourage user uptake. By 2010, online multiplayer games will generate 20.5 percent of total global revenues.

3.1.3 Learning habits

Between the 5 countries results are very different in the learning habits fields. For example, while Croatian, Austrian and Italian students prefer studying alone, British people study best in group or enjoying themselves. This can be inferred also to the fact that students that responded to this survey study different disciplines or are from different faculties.

Regarding technology learning preference, almost homogeneously among the 5 countries three in four of those questioned assert that it is good to have the help from both sides: the human interaction and the technology.

The EASY project aims to make a contribution to the EU's effort to foster the development of new organizational models in higher education, favoring exchange and collaboration also in terms of virtual mobility.

The EASY report "A European Study to monitor e-learning development processes: The On line Campuses and the needs of University Students" analyzes the 46 on-line virtual campuses selected by the EASY Project, and on the e-learning needs of the 1200 students interviewed by the on-line questionnaire in the EASY partners countries (Belgium, Hungary, Italy, Spain, Poland & the United Kingdom).

Of note is the data related to the students who had never had a distance learning experience: they are only 14.93% of the respondents.

Regarding where the students primarily got their information on on-line courses, predominantly this was from the universities themselves (37.6%) followed by the internet (27.6%) and word of mouth (15.8%). Less effective were forms of advertising such as posters, leaflets, TV and Radio (less than 5%). The channels of access to on line course information replicate those of the consolidated channels (university and friends which total more than 52%) and emerging channels such as the internet.

Another interesting finding is the reasons for choosing on-line instruction rather than traditional methods, the main reason being "to save time" (24.6%). The 23.18% of respondents to mGBL questionnaire has the same opinion: "the use of technology saves time, enables fast access to information".

It seems that the courses on-line were chosen for their "novelty" or economic reasons. In fact, reasons related to the possible benefits of the courses, such as "Better Preparation" (13%), "Up to Date Courses" (12%), "Better Results" (5.8%) or to avoid course restrictions were of comparatively minor importance. It is also interesting to note that 2 predicted responses, relating to work needs or work/life balance, were not chosen by anyone.

Table deals with the perceived possibilities of Distance Learning. Students felt that such training enriched their experience (29.8%) or provided greater

access to knowledge in circulation. Of relative importance is the value given to learning specific skills (19.1%) and participating in a work group or network (14.6%). Considering that the students could not respond to all the items, little value is attributed to distance learning as a means of meeting new people (6.7%) this is because the students already belong to a highly socialized context and use internet to maintain contacts already established.

In mGBL survey, students were questioned about mobile phones usage in learning. Of note is the fact that two in five of students seem not to be aware of the mobile learning opportunity and therefore they don't know if it can provide added value to their learning process.

3.2 Evaluation of general first impressions

The questionnaire given at the end of the trial represents the final feedback from the subject users. The first impression of the games is divided into 3 major topics as stated previously:

- Opinion about the games
- Feelings during the game
- Most interesting features of the game, suggestions and comments

Although 137 subjects had undertaken the trial, 78% submitted the questionnaire at the end of the trial (106 subjects).

3.2.1 Opinion about the games

It is interesting to observe the different opinions about the games given by students from different countries; more than a half students from Croatia (60.87%) as well as the large part of the Austrian sample (41.67%) think that the games were interesting, fun and educational. 40% of Great Britain students think that it needs some more work, and half of Slovenian students who have undertaken the trial think that the games were uninteresting and boring. The division of percentages is particularly interesting in Italian sample because it is almost equally divided amongst all categories; the most

interesting is that 22.27% think that games need more work, and 22.73% think that they are uninteresting and boring.

Some typical comments concerning game 1 were that it has too much text, yet little content, makes no sense; it takes a lot of time to read the rules. Game 2 was assessed better in a way that it seems useful in educational sense; it tests the knowledge which made it more interesting and has better graphics. Game 3-Baroque Blog wasn't understood as a game; more as a movie since nothing important (according to the students) happened. They couldn't understand how it works.

3.2.2 Liking of the games

Asked to assess the "liking" of the trailed games, students gave the following answers; 50% of Croatian subjects disliked the game or had no opinion/answer. In Italian sample there was again almost equal division of answers amongst categories (31.82% liked everything, 22% disliked or had no opinion/answer, 27.27% liked the idea, new concept/approach, but none liked speed). 48% of Great Britain students liked the idea, new concept/approach, and 75% of Slovenian students disliked the games or didn't have opinion/give answer.

It is important to state that graphics and speed were poorly rated in all samples.

Comments gave the insight in usefulness of the Game 2 through learning new facts. The decision making in the board game (Game 1) is a very good mixture of education and entertainment. Game 3 is the nice example of using technology as a teaching tool in the classroom.

3.2.3 Disliking of the games

Asked about "disliking" of the games, the subject gave inconsistent answers compared to the previous section; the percentages are not simply "turned around" although it should be expected. According to mention above, 41.67% of Austrian sample stated they disliked the idea, new concept/approach, although the same percentage stated previously they liked the same. It is

important to remind that, although the percentages of filling in the questionnaire at the end of trial vary per countries, 75% of the original Austrian sample responded at the end of trial.

In Italian, Great Britain and Slovenian samples almost 50% of subjects disliked “everything” and in Croatian sample only 10.87% of answers fill in that category, but another 41.30% goes into the poorly defined category of “Disliked (again)/ No opinion/No answer” so it could be, vastly interpreted, summed up to 52.17% disliking for Croatia, 68.18% for Italy, 61% for Great Britain, 75% for Slovenia and 58.33% for Austria.

In overall it represents 62.94% of students in whole sample who didn’t like everything or disliked it/had no opinion/gave no answer.

3.2.4 Feelings during the game

The following terms were offered in order to enable subject the choice of ones aroused during the trial:

- Excited
- Interested
- Curious
- Worried
- Doubtful
- Bored
- Other

From the list of offered terms/feelings, no excitement was stated whatsoever; there was a significant presence of curiosity in all samples (47.83 in Croatian, 50% in Italian, 44% in Great Britain, 25% in Slovenian and 41.67% in Austrian sample) and interest in some samples (Croatian 55%, Austrian 66.67%), but subjects assessed games as boring in also large percentages (e.g. Great Britain 68%, Slovenia 50%, Italy 36.36% and Austria 33.33%).

Therefore it can be assumed that the initial curiosity soon turned into boredom due to some characteristics of the games at the present state of their development.

Comments at the end showed that the language in the Game 1 is too complex. The games, in general, seemed to linear. There is hardly any creativity or major interactions. It is over again stated that there is too much text in the games wrapped in hardly understandable rules. Game 1 is difficult to understand with too much information for a novice. Game 2 is basically a questionnaire and people get tired of it quickly. Game 3 is a movie-lacks the purpose according to students.

3.3 Evaluation of students' ratings for Game 1

Students graded Game 1 regarding its user friendliness, usefulness of information, appeal of the game, satisfactory interface, clearness of the objectives, clearness of playing instructions, game rules, difficulty, activity of a user role, interest in a game, enjoyment, motivation, final judgment of a game, features, suggestions, comments.

46% of subjects considered the game sufficiently user friendly, while 24% found it good adopted to the users.

Usefulness of information provided by the game was sufficient for 39% of subjects while 25% found it to be poor.

The game was sufficiently appealing for 42% of subjects while 32% found it poorly appealing.

The 47% of subjects found the interface satisfactory sufficient, while 23% think it is poor.

32% of subjects think that the objectives of the game were not clear and were explained poorly. 17% even think that clearness of objectives was very poor, but also 17% think it was sufficient enough.

25% of subjects think that instructions for playing game were clear and good explained, but also 25% think that it was poor.

Game rules were sufficient for 25% of subjects, but 24% stated them poor.

Game difficulty was sufficient for 33% of the students, while for 25% it was poor.

25% of subjects think that their role was sufficiently active, while other 25% the activity of their roles in the game was good.

When it comes to interest, 33% of subjects think that the game wasn't so interesting (poor), but 25% think it was good.

42% of the whole sample thinks that enjoyment in the game was poor.

33% of the questioned subjects think that their motivation for the game was sufficient, but 25% of them assessed it as very poor.

Final judgment for the game 1 was poor for 33% of the subjects.

Most interesting features of the game were strategy decision making for 19% of subjects and deal making-acquisitions for 17%.

20% of subjects didn't like the game because it was boring and repetitive, and 19% found it too complex.

29% suggested that new and better game should be made, while 17% think that the existing game should be made more interesting. 11% commented that the game was boring and useless for learning.

3.4 Evaluation of students' ratings for Game 2

Students graded Game 2 regarding its user friendliness, usefulness of information, appeal of the game, satisfactory interface, clearness of the objectives, clearness of playing instructions, game rules, difficulty, activity of a user role, interest in a game, enjoyment, motivation, final judgment of a game, features, suggestions, comments.

32% of subjects found the game 2 sufficiently user friendly. Other opinions were represented with minor percentages.

Useful information provided by the game for 26% of subjects was assessed as good and sufficient for another 23%.

The game was sufficiently appealing for 25% of subjects.

The satisfaction with the interface was sufficient for 24% of subjects while 21% found it good.

Objectives of the game were sufficiently clear for 22% of subjects, while 23% found it good and understandable.

Instructions for playing the game were sufficient for 22% of players and good for 24%.

Game rules were good for 28% of subjects and 19% found it sufficient.

The game was sufficiently difficult for 27% of players, and the difficulty levels were placed good for 23% of subjects.

The role of the players was sufficiently active for 23% of subjects and 24% assesses it as good.

The game was sufficiently interesting for 20% of subjects, but 20% found it also poor and good.

24% enjoyed the game sufficiently, but 19% found the enjoyment poor.

For 29% of subjects motivation for playing the game was sufficient.

Final judgment for Game 2 is that 32% of asked subjects found it sufficient.

The most interesting features of the Game 2 is learning the first aid for 12% of subjects and answering the questions for 11%. 9% disliked the game because it was slow. The same percentage thinks that it has poor graphics. A bit confusing is that "nothing found as the interesting feature" equals "no answer" therefore an amount of potentially useful data is missed since 65% of answers drops into that category.

9% of users suggested using more different questions, or better interface 4%.

8% suggested that something more funnier/better should be made and to definitely avoid man smoking a cigar as a lead in a health quiz. The amount of text should be smaller.

However, users commented Game 2 as the best out of 3 games tested.

3.5 Evaluation of students' ratings for Game 3

Students graded Game 3 regarding it's user friendliness, usefulness of information, appeal of the game, satisfactory interface, clearness of the objectives, clearness of playing instructions, game rules, difficulty, activity of a

user role, interest in a game, enjoyment, motivation, final judgment of a game, features, suggestions, comments.

25% of subjects found the game user friendly on sufficient level.

25% of users found the information provided by the game useful.

33% of subjects found the game appealing.

33% of subjects think that game was interesting good enough.

33% also assessed the enjoyment of the game as good.

When it comes to motivation for the game, 17% of subjects think that their motivation for the game was good.

Final judgment for 25% of subjects is that the game is good, and 17% think it's sufficient.

The most interesting features of the game were finding the information for 8% of subjects and teamwork for 12%. 30% of subjects found the game useless; cannot be played. 17% think of it as a presentation, not even a game.

20% of students suggested to make the game playable/make it a game, and to improve the graphics.

4 Global evaluation of teachers' trials results

The evaluation of teachers' trial results is based on answers and comments given by 13 teachers in the questionnaire filled in at the end of game testing according to D6.2 First User Trial Report. Responding teachers represented the following fields of expertise: engineering and informatics, education, computer science and motivational program for drop outs.

Questionnaire consisted of teachers' first impressions, pedagogical aspects of the game, organizational aspects of the game, communication aspects of the game and skill aspects of the game. They were asked for their opinion about the game, what did they like or dislike, were their prior expectations met, were the stated objectives met and were their individually expected goals reached. Teachers were also asked to state their feelings during the play and to rate their answers to the questions referring to individual games. They chose among the following responses: excellent, good, sufficient, poor, and very poor. The responses "excellent, good and sufficient" are considered as different levels of positive comments, and "poor" as well as "very poor" as negative. There were three games tested: Game 1 "On the edge", Game 2 "Ahead of the game" and Game 3 "Baroque Blog".

4.1 Evaluation of general first impressions

Teachers' first general impressions of the games were mostly favorable. The key word describing the prototypes was "interesting" in the sense that they offered a dynamic way of learning that is not traditional, a way of reaching young people, of motivating them for learning and a way of making a school subject more interesting.

Only one opinion can be described as negative due to insufficient user friendliness, lack of effectiveness and affect and too complicated to enable playing without reading help and instructions first.

When teachers were asked directly what they liked their responses were again favorable mentioning dynamics, graphics and supporting role in the educational process. Game 3 was mentioned twice as a likeable game with an interesting comment that it was centered on actually doing something with your mobile and couldn't have been implemented by using either paper or computer. One respondent didn't like anything.

When directly asked what they disliked teachers commented the graphics in Game 2 as irrelevant to subject matter. Some had difficulties with getting used to the idea that mobile phones can be used for educational purposes. Game 1 was mentioned once as unsuitable for small screen. Other comments included limited interaction, abundance of text, inconsistent use of buttons and insufficient instructions.

In general, 71% of teachers stated that their expectations were met in the user trials. The prevailing feeling during the test was curiosity (77%) and the least occurring feeling was boredom (8%).

4.2 Evaluation of teachers' ratings for Game 1

Teachers graded game 1 according to their opinion regarding its user friendliness, usefulness of information, appeal, satisfactory interface, clearness of objectives, clearness of instruction for playing, being comfortable with game rules, difficulty, role activeness, whether it was interesting, enjoyable, what was the level of motivation and gave final judgment of the game.

67% of teachers were of opinion that user friendliness is sufficient at least with majority deciding for the option good. 8% believed it was excellent and 8% very poor.

84% of teachers thought that usefulness of information was sufficient (48%), excellent (25%) or good (17%). The remaining 16% were equally divided between poor and very poor.

Appeal of the game was considered good by 40% and sufficient by 30%. None of the respondents opted for very poor and 10% opted for excellent.

If we consider marks poor and very poor as negative while sufficient, good and excellent as supporting marks, half of the respondents graded interface as negative, 17% of them as very poor. In the other half marks were distributed among good (25%), sufficient (17%) and excellent (8%).

Marks on the clearness of objectives were equally distributed among good (25%), sufficient (25%) and poor (25%). 17% believed it was excellent and 8% it was very poor.

For playing instructions the highest percentage was given to option poor (34%). Remaining were equally distributed between good and sufficient (25% each) and between excellent and very poor (8% each).

34% believed game rules were poor but on the there hand 33% opted for good and 17% for excellent. Sufficient and very poor carry 8% each.

Only 9% were of the opinion that game difficulty is poor and no one that it is very poor. 37% said it was sufficient, 36% that it was good and 18% that it was excellent.

All respondents felt that their role was active during play. 42% opted for sufficient, 50% for good and 8% for excellent.

A large majority of the respondents (good 51%, sufficient 33%, excellent 8%) felt that the game was interesting. 8% thought it was poor and none thought it was very poor.

Most of them enjoyed the game (good 42%, excellent and sufficient 25% each) whereas 8% stated that their enjoyment during play was very poor. Teachers were highly motivated during play. 51% opted for excellent, 25% for good and 8% for sufficient. The remaining 16% were equally distributed between poor and very poor.

In final judgment of the game two thirds of respondents (67%) said the game was good and 25% that it was sufficient. None decided for excellent or poor.

As we can see, grade 4 was the highest and given to motivation. Other grades higher than 3.5 were 3.67, 3.58 and 3.75 given to active role, interesting characteristic and enjoyment, respectively. Appeal, interface and instructions

received the lowest grades (below 3) in the trial. Interface received the lowest grade 2.75.

According to standard deviation data teachers were most unified answering the question about role activeness (0.62) and least unified with giving opinion about interface (1.23).

The outstanding comment regarding the most interesting features in the game was multi-player option which was commended by 4 teachers. Others felt the general idea, strategic playing, graphics and creativity were interesting.

Comments to what they hadn't like in the game correspond to the grades mentioned above: unclear instructions, inappropriate interface, only some options were available, the game is linear and results are unpredictable.

4.3 Evaluation of teachers' ratings for Game 2

Teachers graded game 2 according to their opinion regarding its user friendliness, usefulness of information, appeal, interface, clearness of objectives, clearness of instruction for playing, being comfortable with game rules, difficulty, role activeness, whether it was interesting, enjoyable, what was the level of motivation and gave final judgment of the game.

90% of teachers were of opinion that user friendliness is sufficient at least with majority deciding for the option excellent (40%). 10% believed it was very poor.

80% of teachers thought that usefulness of information was either excellent (30%), or good (30%) or sufficient (20%). The remaining 20% were equally divided between poor and very poor.

Appeal of the game was considered good by 40% and excellent by 20%. 10% opted for sufficient and 10% for poor. 20% opted for very poor.

60% considered interface good, 20% either sufficient and 20% poor. The remaining 10% decided for very poor and none for excellent.

Marks on the clearness of objectives were evenly distributed between excellent (30%) and good (30%) and between sufficient (10%) and very poor (10%).

For playing instructions the highest percentage was given to option good (50%). Excellent followed with 30%. The remaining 20% were equally distributed between sufficient and poor. None opted for very poor.

40% believed game rules were good and 30% they were excellent. Sufficient carried 20% and very poor 10%.

30% were of the opinion that game difficulty is good. 20% were allocated to excellent, sufficient and poor. For very poor opted 10%.

80% of teachers felt that the activeness of their role was good (60%) or sufficient (20%). The remaining 20% were all allocated to very poor.

A majority of the respondents (sufficient 30%, excellent 20%, good 20%), felt that the game was interesting. 10% thought it was poor and 20% thought it was very poor.

Most of them enjoyed the game (good 40%, sufficient 30%, excellent 25%) whereas 10% stated that their enjoyment during play was very poor.

Teachers were motivated during play. 50% opted for good, 30% for sufficient and 10% for excellent. The remaining 10% were allocated to very poor.

In final judgment of the game 30% of respondents said the game was good, 30% that it was poor and 20% that it was excellent. 10% decided for sufficient and 10% for poor.

As we can see, grade 4 was the highest and given to clear playing instructions. Other grades higher than 3.5 were 3.80, 3.80 and 3.60 given to user friendliness, game rules and enjoyment, respectively. Game 2 received 1 grade under 3. The lowest grade was given to game appeal (2.83).

According to standard deviation data teachers were most unified answering the question about playing instructions (0.76) and least unified with giving opinion about usefulness of information (1.33).

The outstanding comment regarding the most interesting features in the game was that it enables learning new things and taking decisions. Some teachers appreciated graphics and idea of a game and seemed ok with instructions, answers and scoring.

Comments to what they hadn't like in the game mainly referred to the quiz characteristics of the game or linearity of the game. Teachers who were not involved in medicine would have preferred other content. The graphics style was not found suitable to the content by one teacher. Interface was mentioned twice and missing signal notifying that time is running out once.

4.4 Evaluation of teachers' ratings for Game 3

Teachers graded game 3 according to their opinion regarding its user friendliness, usefulness of information, appeal, whether it was interesting, enjoyable, what was the level of motivation and gave final judgment of the game.

90% of teachers were of opinion that user friendliness is sufficient at least with majority deciding for the option good (40%). 10% believed it was very poor.

91% of teachers thought that usefulness of information was either good (46%), or excellent (27%) or sufficient (20%). The remaining 9% were allocated to very poor.

Appeal of the game was considered mostly sufficient by 37%, good by 27% and excellent also by 27%. None of the respondents opted for very poor and 9% opted for poor.

A majority of the respondents (sufficient 33%, excellent 33%, good 17%), felt that the game was interesting. 17% thought it was poor and none thought it was very poor.

Most of them enjoyed the game (excellent, good and sufficient were given 25% each) whereas 17% stated that their enjoyment during play poor and 8% very poor.

Most of the teachers were at least sufficiently motivated during play (34%). Excellent and good motivation was equally noted by 25% respectively and poor or very poor by 8% each.

In final judgment of the game half of respondents said the game sufficient, 25% that it was excellent and 17% that it was good. None decided for very poor.

As we can see, grade 3.82 was the highest and given to usefulness of information. All remaining grades were higher than 3.5 were but for enjoyment factor which was graded slightly below (3.42)

According to standard deviation data teachers were most unified answering the question about game appeal (0.83) and least unified with giving opinion about enjoyment (1.15).

Teachers very much liked the idea of the game building team work and students making active contribution to the usual classes. The activity had a component of adventure but it still is more an activity than a game.

Comments to what they hadn't like in the game mainly referred to the opinion that it cannot be used in all learning situations and that the prototype or a demo was not a game as such. They could only observe in a passive way and advised that the supporting material for this activity should be well prepared.

4.5 Pedagogical concepts of the games: a reasoned summarise of teachers' and guidance practitioners' comments after the first users' trials

4.5.1 Context

During First User Trials teachers and guidance practitioners were asked to give their comments on the positive and negative features of first prototypes of three games: "*On the edge*", "*Ahead of the game*" and "*Baroque blog*".

In this chapter we included their comments on all features of the games as we consider that all aspects contribute to the pedagogical "success" of a learning and guidance tool, being them interconnected and contributing to transfer contents, i.e. influencing the potentialities of contents to be "assimilated" and "reflected upon" better in one way than in another one by pupils.

We also have to affirm beforehand that **some times comments are not easy to understand** as the questionnaires were available in English in this first testing phase and that could have hindered the comprehension and right explanation of what teachers and guidance practitioners wanted to say.

Nevertheless, interesting feedbacks have been given and they have to be taken into consideration in the “adaptation” of these games as these users (teachers and guidance practitioners) will be those ones who will use these devices for educational and guidance purposes.

Only few users considered these games as non-effective and no user-friendly at all and most users considered these suggestions according to an “ameliorative” approach, being these games only in a first prototype phase.

A more general comment was also that to pay attention to the fact that economic conditions do not hint the use of these devices by students with little economic possibilities.

In the next sections we collected users’ comments under the same categories according to the positive and negative perception issues of teachers and guidance practitioners that in this chapter will be defined with a general “users”. This division can help to make any comparative analysis among them.

4.5.2 Positive perception of the games

Layout

Most users (in this case teachers and guidance practitioners – we will use the term “users” within this text when referring to them) considered the scenarios of these games quite interesting and amusing even though, being these games still prototypes, some faults were detected in the layout.

Instructions on the usage of these games were particularly appreciated by users even though in some cases they asked for more details in that.

With reference to the “Baroque blog” game users liked its graphics and considered it a very important feature in order to support tuition.

Interaction

Users liked the high level of dynamic interaction of some of these games and considered a good chance to use them within their classes or guidance pathways.

When commenting the game "On the Edge" users consider it conceptually good and in particular they appreciate the multiplayer possibility.

Learning and guidance potentialities

Some users took into consideration the learning and guidance potentialities of these games to be exploited "within traditional" learning or guidance contexts (i.e. classes) but that is only the first users trials. That could be an interesting indicator to be analyzed in any specific question in future tests.

The big potentialities to reach young people and to "transfer" learning and guidance contents have been acknowledged by quite all users even though some of them admitted to have not so much experience in the field of mobile games for learning and guidance purposes. The use of games was acknowledged also to have good learning potentialities in addition to the communication ones. That is an important aspect nevertheless as these games will have to be used for learning and guidance purposes by this target users, not having big experience in this field (see also results coming from WP2 users' interviews). A good perception of these games potentialities for learning and guidance purposes is an added value as, very simply, the game will be used, if it is perceived as a valid learning and guidance tool.

With reference to the "Baroque blog" game users considered its graphics helping the learning and guidance process.

An additional positive aspect is that students in the game are active, they have to cooperate and this is a device to test their social skills.

They considered it as funny and an added value is that the team activity contributes to warm the atmosphere better than a simple quiz.

In addition, this game offers the view of a real class situation and this is an added value for learning purposes.

With reference to "Ahead of the game" it is the game with the most knowledge contents while comprehension is higher in the game "On the Edge".

Teachers and guidance practitioners considered these games application particularly interesting in a blended context, i.e. within traditional classes or in guidance pathways. They considered these devices important in order to change both the teaching standard methodological approach of teachers and guidance practitioners and the learning way of students.

4.5.3 Negative perception of the games

Layout

Most users did not like very much the image of the businessman in the prototype "Ahead of the game". In addition, its graphic was considered no relevant to the subject and often the cigar of the boss or the fat cat in the game are not considered a good example for students, especially if the game content is focused on e-health. Some users considered quite hard to use the game on a mobile screen.

Some users considered quite hard to use the game "On the edge" even on a PC, so they could not imagine its use of the screen of a mobile.

Some of them suggested an accurate overhaul of the game interface, taking into consideration its learning purposes.

Another comment was on the use of texts to transfer contents, especially on the use of too many words within this prototype (a suggestion was to follow the Pokemon game). Some difficulty in using "buttons" was also detected.

For some users a simpler interface would be better and also the game speed and sounds are very important.

Interaction

Some users considered very important to have sounds and not only images. In addition, they considered the interaction possibilities in this first prototype phase are very limited.

Users considered these games in general should be more interactive with fewer texts to be read.

Some users considered the "Baroque blog" having no interaction possibilities, being only a demo version.

With reference to the game "Ahead of a game" most users considered it could better take advantage from the possibilities offered by mobile technologies and not having it just as a quiz.

Learning and guidance potentialities

Sometimes teachers and guidance practitioners are skeptic on potentialities these devices can have in the educational and guidance field but we also have to consider that not having the possibility to "visualize" these games in their complete and final framework (so, not as prototypes) contribute to this pre-assumption. Nevertheless, there is a sort of curiosity also in skeptical users who suggest envisaging a "future device" to support these games.

Some users consider it is quite difficult to evaluate the learning potentialities of these games, as they are not complete anyway. Nevertheless, they consider these games a good way both to evaluate students' competences (skills, knowledge and attitudes), for example is a testing phase, and to teach new contents.

Some of them considered that the same contents could be transferred through the most traditional use of paper or even on a PC. This type of consideration was done especially when testing the game "Baroque blog".

In this game some users suggested more supporting material in order to improve its learning and guidance potentialities. In addition, some users considered this game not fit for all contents and for all students according to their personal features.

Some of them considered it as not a real game but as an “observing game” even though this concept should be clarified, as it is not clear the exact meaning. According to the suggestion given to this critical item, probably the meaning is that the player should be more involved in the game and not be an “external observer”.

With reference to the game “On the edge” some of them did not catch very well the aim.

Some of them pointed out that not all options are available and perhaps additional choices (fields) could be added.

Finally, instructions could be better defined.

5 General analysis and cross comparisons

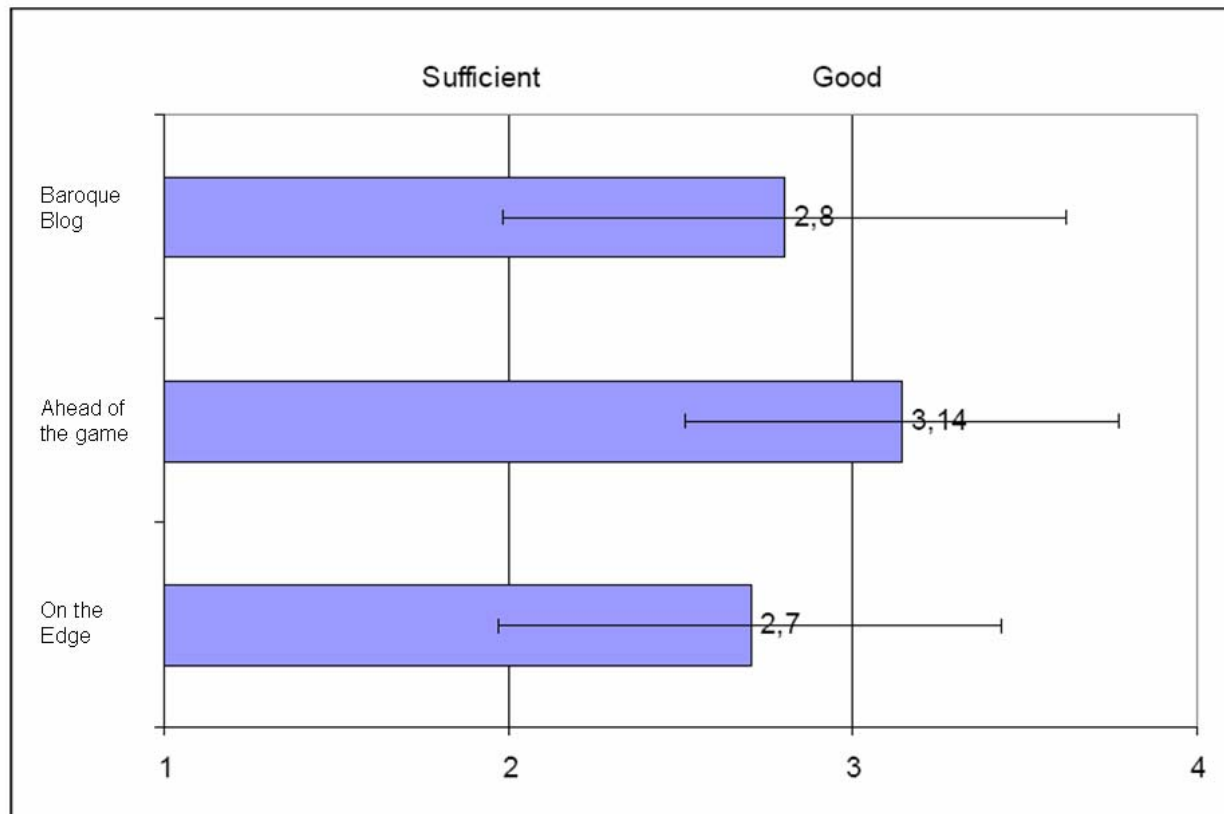
In this analysis, “open” questions can be a rich source of information. During user trials, students have been motivated enough to fill in the questions fully and they gave us interesting feedback.

The idea of mobile games that supply also useful information has been appreciated from the most of the students. However analyzing the responses, the 3 games appeared too little “exciting”. Although interviewed students were aware that they were evaluating educational games, they expected to have more “fun” and interaction in the games. During the user trials the students’ attention was more concentrated on the playable and graphic aspects, rather than by the educational aim of the games.

Regarding “On the edge” game, for the most of the students it seemed incomplete yet: there were no objective and too few contents. The initial instructions, too long and boring, annoyed and distracted the interviewed from the playing successive phase.

“Ahead of the game”, appreciated for the graphic appealing, appeared to a lot just a questionnaire, for someone potentially useful for studying, for others boring and frustrating to answer.

Analyzing the responses about the third game, “Baroque blog”, it can be inferred that there was a general misunderstanding: the most of respondents thought that the game was the movie, which should be just a presentation of the basic concept of “Baroque blog”. The expressed impressions are biased by this misunderstanding, due probably to insufficient information/preparation given to the users before the trial.



Comparison of the games

Comparing the ranking of the different aspects of the game, the most appreciated game was absolutely “Ahead of the game”, the most completed in this phase of the mGBL project and with a more winning graphic.

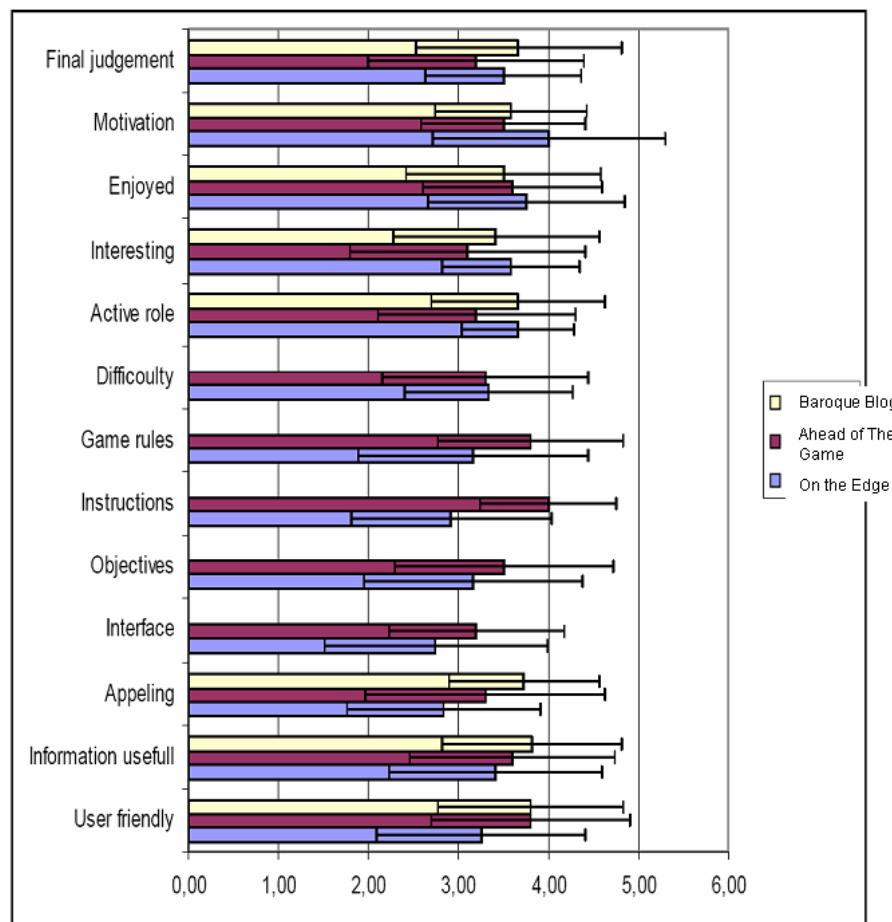
The other 2 games paid the fact that they were not completed or that the respondents based their valuation only on the playful side of the games and they wouldn’t/couldn’t give any judgment on the games’ concept. This is can be referred also to a wrong or insufficient design of the user trials.

Some nice results emerge from the study of teachers’ responses: unlike the interviewed students, teachers didn’t base their valuation on the first glance (graphics, sounds, ect...) that was neglected in this phase of implementation.

The concept of mobile games with an educational aim likes and stimulates curiousness and interest.

About "On the edge" game, it was appreciated the strategic side and the possibility to interact with the other players, although the game's objective misses. In "Ahead of the game", the easy but winning interface likes mostly, even if some teachers expected more than a simple quiz game.

Regarding "Baroque blog", as for the most of interviewed students, some didn't realize that it was just a video demo of the game. Then the responses were influenced by this misunderstanding.



Comparison of the games

Comparing the 3 games results, "Ahead of the game" was appreciated mostly for interface and playing objective, "On the Edge" for active role and motivation in playing, "Baroque blog" for the interaction offered within the class.

According to the general description of mGBL project, success criteria will include “enjoyment of use of mGBL prototype by at least 75% users involved in trials”. If we consider the appraisal given by students and teachers to the three games (sum of favorable answers at the questions “Have you enjoyed the game?” and “Final judgment of the game”), these criteria are not fully satisfied yet in this early phase of project.

In the tables below, the positive judgments to the games, if we consider marks poor and very poor as negative while sufficient, good and excellent as supporting marks.

“Have you enjoyed the game?”

	On the Edge	Ahead of game	Baroque Blog
<i>Students</i>	25%	42%	33%
<i>Teachers</i>	92%	90%	75%

“Final judgment of the game”

	On the Edge	Ahead of game	Baroque Blog
<i>Students</i>	34%	52%	42%
<i>Teachers</i>	92%	60%	92%

6 Suggestions for improvement of the games

The following suggestions for improvement of mobile games arise from students' and teachers' comments during the first user trials.

General:

- improve instructions (clear menu structure)
- offer help only where it's needed
- make sure that user can play directly without using help (conserve the user's time)
- decrease the amount of text
- make sure that the game interaction is not limited
- improve some faults in game design
- adapt the games to small screen
- add sound
- make them for future use

Suggestions for Game 1 ("On the Edge"):

- work on interface, make it simpler
- define the playing objective
- provide useful feedback on own current status and the status of other players
- add more levels to the game, make it more complex
- include sound features

Suggestions for Game 2 ("Ahead of the game"):

- loose the graphic of the businessman with cigar
- take advantage of the multimedia features of the phone

- be variable to keep players interested (bonus and special features)
- don't just test knowledge, develop skills
- include signals denoting running out of time

Suggestions for Game 3 ("Baroque Blog"):

- provide more details to the game concept
- involve the user
- provide suitable phones to the students

7 Suggestions for next trials

7.1 Methodology

7.1.1 Introducing User Trials

According to Carol M. Barnum the definition of user trials and testing is:

"The process of learning from users about a product's usability by observing them using the product, one aspect of usability engineering, is called usability testing"

Another scientist, Jeffrey Rubin, says that *"Usability Testing employs techniques to collect empirical data while observing representative end users using the product to perform representative tasks. The main goal is to expose usability deficiencies and modify the product."*

With the help of user trials developer are able to test **usefulness, effectiveness, learnability** and **attitude (likeability)**.

"In user trials a product is tested by "real users" trying out the product in a relatively controlled or experimental setting, where they are given a standardized set of tasks to perform. The result can be a "problem list" which contains valuable information for designers regarding the potential for improving the usability of a product."

(Definition User Trial:

http://ec.europa.eu/information_society/policy/accessibility/deploy/dfa/a_documents/userfit/4-11Usertrials.pdf)

7.1.2 Different Types of Usability Tests

There are miscellaneous types of usability tests (exploratory, assessment, validation, comparison). The difference between the first three is that they are administered at different point in the product development cycle. The fourth test can be an integral part of any of the other three tests, and is not associated with any specific life cycle phase.

7.1.2.1 Exploratory Test

The Exploratory Test is conducted quite early in the development cycle, when the product is still in the preliminary stages of being defined and designed. The functional specifications and early models are in the development phase. A further goal could be the inspection of the specifications at the beginning of the design phase. In this early stage it's common to use a **paper prototype test** or **Wizard of Oz technique**. According to Jakob Nielsen the advantage is that *"with a paper prototype, you can user test early design ideas at an extremely low cost. Doing so lets you fix usability problems before you waste money implementing something that doesn't work."*

(Jakob Nielsen, <http://www.useit.com/alertbox/20030414.html>)

"The Wizard of Oz technique enables unimplemented technology to be evaluated by using a human to simulate the response of a system". One of the most essential goals is to *"test the interaction of a device before building a functional model"*. (<http://www.usabilitynet.org/tools/wizard.htm>)

7.1.2.2 Assessment Test

The assessment test is probably the most typical type of usability test which is conducted either early or midway into the product development cycle, usually after the specifications, the fundamental or high-level design has been established and (less or more) functionality is implemented into the prototype version.

The purpose is to expand the findings of the exploratory test by evaluating the effectiveness of realistic tasks performed by the test user - the effectiveness of preliminary design concepts will be tested. Thus, the developers are able to identify specific usability deficiencies and their causes for the most common tasks.

7.1.2.3 Validation Test

The validation test (also called verification test) is conducted late in the development cycle. The purpose of this test is to verify that the product corresponds to certain standards (e.g. usability standards, project-related

early defined performance standard, internal company standards). The developers investigate how fast and effectively a user performs certain tasks. Example: A setup procedure for a software package has to work well and must be conducted within 15 minutes with no more than one error.

In many cases the results won't be implemented in the product before its release (because of a lack of time), but in the next versions of the products. Moreover, the support team can be trained and public-relation responses can be prepared.

7.1.2.4 Comparison Test

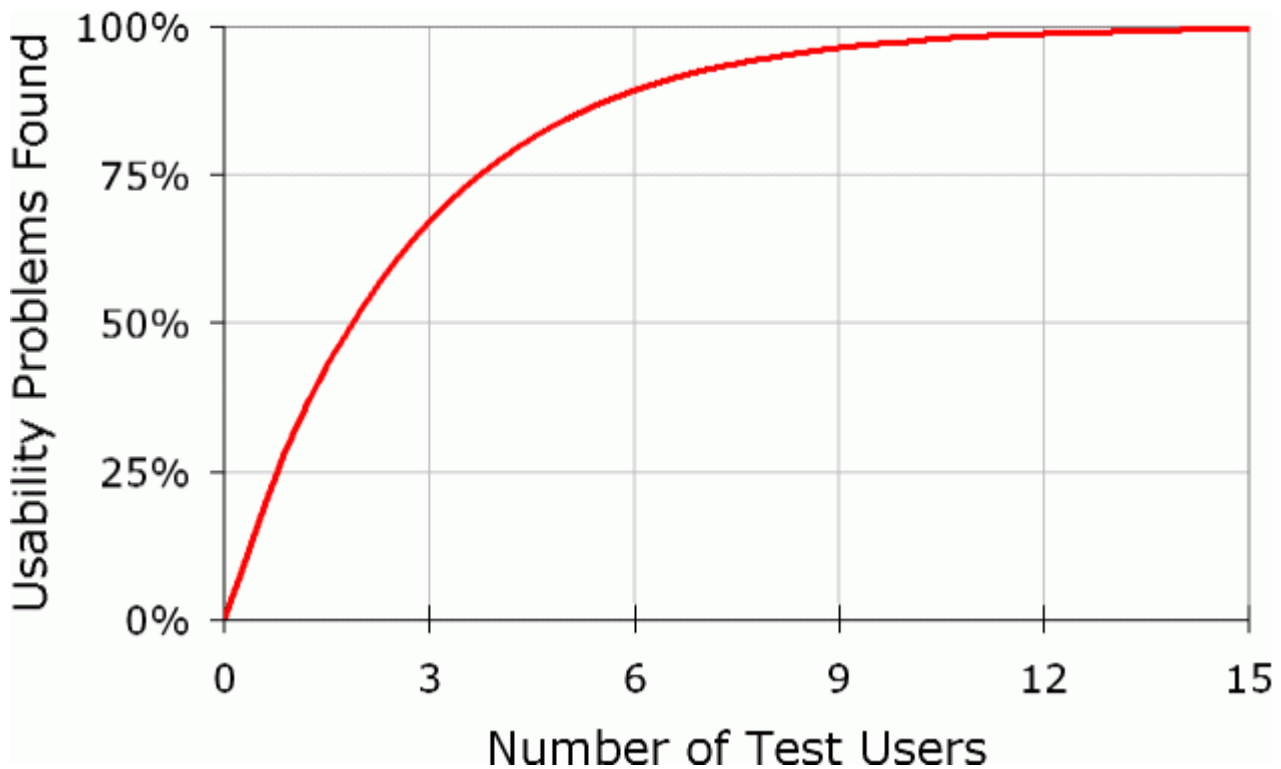
A comparison test can be conducted at every stage in the product development cycle. At the beginning it could be used to compare several radically different interface styles, and in the middle (when you have a graphical interface) you can evaluate single elements like buttons. Finally, the ultimate prototype can be compared to the competitor's product.

7.1.2.5 How many users do you need to get the most problems?

According to Jakob Nielsen the number of usability problems found in a usability test with n users is:

$$N(1-(1-L)^n)$$

" N is the total number of usability problems in the design and L is the proportion of usability problems discovered while testing a single user. The typical value of L is 31%, averaged across a large number of projects we studied."



“The best results come from testing no more than 5 users”

Nielsen says:

“The most striking truth of the curve is that zero users give zero insights.

As soon as you collect data from a single test user, your insights shoot up and you have already learned almost a third of all there is to know about the usability of the design. The difference between zero and even a little bit of data is astounding.

When you test the second user, you will discover that this person does some of the same things as the first user, so there is some overlap in what you learn. People are definitely different, so there will also be something new that the second user does that you did not observe with the first user. So the second user adds some amount of new insight, but not nearly as much as the first user did.

The third user will do many things that you already observed with the first user or with the second user and even some things that you have already seen twice. Plus, of course, the third user will generate a small amount of new data, even if not as much as the first and the second user did.

As you add more and more users, you learn less and less because you will keep seeing the same things again and again. There is no real need to keep observing the same thing multiple times, and you will be very motivated to go back to the drawing board and redesign the site to eliminate the usability problems.

After the fifth user, you are wasting your time by observing the same findings repeatedly but not learning much new.”

7.2 Application to mGBL – Project

7.2.1 Second User Trials

For the second user trial the most appropriate type is the assessment test so the effectiveness of preliminary design concepts can be tested. Thus, the specific usability deficiencies and their causes for the most common tasks will be identified.

7.3 Schedule for User Trials

(Accepted theory for Usability Testing, according to Carol M. Barnum and Jeffrey Rubin)

A Usability Test consists of several stages: Planning, Preparing, Conducting and Analysis/Reporting.

7.3.1 Planning for Testing

- Establish a team (team structure, size), who will conduct the test

- Defining product issues and the audience. The time used for the test should be maximum one hour. The test should be done under controlled conditions not as an online test.
- Define the type of test (think-out-loud, co-discovery, exploratory)
- Define the number of the users that have to perform the test.
- Setting goals and measurements for the test. The specific usability deficiencies and their causes for the most common tasks should be identified. All project partners should participate in the goals and measurements choice.
- Selecting the tasks to include in the test. A task list for each game has to be prepared. All project partners should participate in the selection.
- Determining how to categorize the results. The results have to be arranged according to their severity.
- Establishing the User Profile (range of age, education...)
- Writing the test plan: The test plan should contain: purpose, problem statements, user profile, methodology (test procedure), task list and test environment and equipment requirements. Test plan should be included in the final report.

7.3.2 *Preparing for Testing*

- Preparation of Test materials*
- Defining team members' roles (moderator, logger, camera operator, etc.)
- Recruitment of the users with the help of screening questionnaires
- Conducting the walkthrough (internal "test" of the user test)
- Conducting the pilot (test with one external person)

***Test materials**

⇒ Screening Questionnaire (for selecting the participants)

- ⇒ Background Questionnaire / further screening for arrival (provides historical information about the participants)
- ⇒ Orientation Script: for the moderator, a kind of moderator guide
- ⇒ Data collection instruments [data loggers] (A form where the data logger can fill e.g. the needed time for a certain task, the number of errors, when an error happened, the "quotations"/comments of participants)
- ⇒ Nondisclosure agreement and tape consent form
- ⇒ Task Scenarios: are expanded versions of the task list
- ⇒ Post-test Questionnaire at the completion of the test (Written Questionnaire)
- ⇒ Debriefing topics guide (List of general topics to discuss)

7.3.3 Conducting the Test

- Greeting and setting the participant at ease
- Administer forms
- Briefing the participant on the process (explaining "thing-aloud" protocol)
- Being an unbiased briefer
- Collecting and organizing data during the test (written notes, video)
- Debriefing the participant
- Collating data into findings (interpreting qualitative data, recording of positive findings and problems)

7.3.4 Analysis of Results

The findings of the user test have to be analyzed. It is important to determine the:

1. Causes
2. Scope
3. Severity of problems

In order to that, there are many charts – the most common is the ranking: “cosmetic-minor-medium-major-catastrophic”.

It is crucial not to forget the positive findings!

Moreover, it is important to make recommendations.

7.4 Reporting the Results

The final report should consist of a written report, an oral report (presentation for visual support) and if the test was recorded, a video-highlights tape.

The written report should have the following parts (in correct order):

1. Cover memo/letter
2. Executive summary
3. Introduction
4. Methodology
5. Results
6. Recommendations/actions
7. Appendices

[**Parts of the written report in detail:** goals of test, problems and concerns of the test, user profile, number of users, tasks and methods, test plan, procedures, length of sessions, date-time-place of test, type of test (think-out-loud, co-discovery, exploratory), system specifications that have an input on the results like connection speed and web browser, state of product (prototype, partial product, full product), measurement instruments (qualitative, quantitative), severity scale, Charting of performance, subjective measures (percentage of participants able to find....) Summary of findings (global findings, local findings), Explanation of usability problems with detailed recommendations]

8 Conclusions

Focal points resulting from these first user trials are described here.

Game appreciation by users was generally rather good, with better appreciation by the teachers. This is important: teachers have better experience and maturity and many of their comments and answers showed they understood better both the potentialities of the games and the fact that they had to judge the concept and the structures of the proposed games; so teachers' answers can be taken as more relevant for this evaluation phase.

Several of the students' answers on the other way reveal that at least some of the users didn't understand fully the aim of the trials. Some comments regarding the poor graphics of game 1 or the consideration that game 3 was just a "movie" are significant examples of such a misunderstanding.

During these first user trials, game 2 "Ahead of the game" was the one in the most advanced phase, game 1 "On the edge" was still in prototypal phase and game 3 "Baroque blog" was just an idea to be realized. Some students weren't capable to understand this and gave their evaluation to the games' interface rather than to the games' concept. This probably led to a negative bias in students' evaluation.

Nevertheless both students' and teachers' answers, in particularly the "open ended" ones, gave precious feedback to better understand their needs. Even the "wrong" answers by users who judged more the interface (graphics, text, instructions) rather than game idea gave us anyway useful hints regarding what is expectancy from students regarding game interface.

Anyway it must be noted that the several answers and comments by users that were somehow "off topic" are revealers of some faults in user trials management. **All participating users should have been thoroughly briefed regarding trials objectives, games stage of development and**

mGBL project aims. The fact that users gave “wrong” or “not pertinent” answers is really a fault of those who had to instruct users during the trial.

Some parts of the questionnaires contained repetitive, little pertinent and not fully clear questions and the total length of questionnaires to be filled probably exceeded the “attention span time limit” of the users.

The number of participants to the trials was probably not thoughtfully chosen, since they were too many to allow an accurate “controlled environment trial” in the short time at disposal to conduct the trials.

Chapter 6 of this report contains a series of analysis and suggestions that will have to be considered to obtain the maximum of information from next user trials and usability testing.

In short, methodical considerations must be done in the following areas:

- **Collecting background information** to backup the interpretations and conclusions;
- Considering the **choices of the evaluators**;
- **Defining success criteria** for evaluation;
- **Documenting of evaluation**, describing the methods and justifying them;
- **Defining the number of the users** that have to perform the test and the **criteria for the selection of participants**;
- **Conducting the test under controlled conditions** (not as on-line test) and **collecting systematic records during the test** (e.g.: written notes, videos...).

- **Sources:**

Carol M Barnum, Usability Testing and Research

Jeffrey Rubin, Handbook of usability testing - How to plan, design, and conduct effective tests

Jakob Nielsen, <http://www.useit.com/alertbox/20030414.html>

Jakob Nielsen, <http://www.useit.com/alertbox/20000319.html>

Definition User Trial: <http://www.stakes.fi/include/usertria.pdf>

Wizard of Oz-Method: <http://www.usabilitynet.org/tools/wizard.htm>

NPD Wireless Bulletin: http://wireless.npd.com/bulletin_mobilegaming_032007.html

EASY report: http://www.easy-elearning.net/report_home.php

Eurobarometer report:

http://ec.europa.eu/public_opinion/archives/ebs/ebs_274_en.pdf

Nokia survey: http://press.nokia.com/PR/200606/1054096_5.html