



Around an Inspiring Virtual Learning World in Eighty Days
Seventh Framework Programme, Objective ICT-2007.4.1, Contract Nr. 215918

Evaluation Results of 80Days' First Implementation Phase

SUMMARY OF PHASE 1 EVALUATION REPORT

OVERVIEW

The primary goal of evaluating the first game prototype of 80Days (a.k.a. Demonstrator 1) is to evaluate usability and user experience and to validate learning efficacy resulting from the gameplay. A **mixed-method approach** aiming to gather both quantitative and qualitative data has been adopted. This evaluation methodology is based on triangulation of measures covering various dimensions: technical, affective, cognitive, and pedagogical.

There were two major phases of evaluation. In the first phase, an initial game design concept was developed prior to implementation. A survey with the aim of assessing the acceptance towards the game design was conducted with children in Germany and England (Study 1). In the second phase, Demonstrator 1 was produced. Expert reviews with *game usability* [1] and *game playability heuristics* [2] on the beta version were performed to eliminate major usability problems prior to testing it with children. User tests primarily in the form of observations and questionnaires were then conducted (Study 2). Results of Study 1 and Study 2 indicate significant gender differences in a range of variables. To further explore the issue of gender differences, two single-gender focus groups were conducted (Study 3).

METHOD AND PROCEDURE

Study 1: Survey of Initial Game Concepts

Design and Participants. The questionnaire consists of two major parts. Part A contains questions on the respondent's gender, age, gameplay habit, gametype preference, and affinity for geography. Part B addresses different aspects of the game. First a synopsis of the game story is presented. Then the following questions are presented: perceived interestingness of alien stories in general and of the game story in particular; improvement suggestions; identification with the story's main player characters; preference of non-player characters; intention to play the game in the future. Two samples from Germany (N = 139; girl:boy = 78:61) and England (N = 142; girl:boy = 59:83) were involved in the survey. They were school children aged between 11 and 14, the target groups of the game.

Study 2: User Tests of the Executable Prototype

Design and Participants: User tests involved children of target age groups in British schools and international schools in Austria, where English is the main instructional medium. The user interface of the game is in English. The user tests were conducted in groups of various sizes, ranging from 4 to 14, in the rooms within the respective school premises. Each participant was allocated to one computer where the game was installed and played it on an individual basis. One or two researchers were present during the tests to provide help and observe the participants' performance and behaviours. The arrangement of the test session is summarized in Table 1. The instruments listed therein have been developed by the project's research team.

Table 1. Overview of the arrangement of a user test session

Activity	Objective and Instrument
Introduction	Describe the aim of the evaluation tests and instruct how to operate the laptops and headsets
Fill in the Background Questionnaire	Items: Identifier (ID), gender, age, gameplay frequencies, gametype preference, affinity for geography, subject grades, early involvement, and expectation
Fill in Pre-test Assessment of Learning Questionnaire (Pre-ALQ)	16 domain-specific questions, open and close-ended, are based on the content of the game.
View Tutorial	6 open- and close-ended questions about the usefulness and usability of the tutorial material and presentation

Total Pre-Gameplay time: ~ 30 minutes	
Play each of the four micro-missions and fill in “After Mission Questionnaire” (AMQ) right away	Questions of AMQ are adapted to the content of the respective micro-mission. Close- and open-ended questions on ease of use, understandability, enjoyability, usefulness of hint, and game duration are included.
Total Gameplay time: ~52 minutes	
Fill in the Post-test Assessment of Learning Questionnaire (Post-ALQ)	The same questionnaire used for Pre-test. The rationale is to assess whether the children’s knowledge of the geographical concepts covered in the game can be enhanced after playing it.
Usability and User Experience Evaluation of the Game Features	It consists of six sections with each of them focusing on different aspects of the game: General experience with the game; Experience with non player characters; Cognitive load; Game components; Game features; Improvement suggestions;
Debriefing	Summarize the activities of the test session and thank the participants
Total Post-Gameplay time: ~33 minutes	

Study 3: Focus Groups: Procedure and Participants

Prior to taking part in focus groups, participants were asked to play through the whole game without being required to fill in any questionnaire except the one for background data. The key data collection steps include:

- Game recall exercise: Each participant was given a stack of Post-it notes and asked to write down whatever they could remember about the game.
- Sharing game recollections: Participants were asked to stick their notes, based on their own judgment, to three sheets labeled as “positive’, ‘negative’ and ‘neutral’
- Guided discussions on different aspects: Gameplay, game characters, game story, and learning content.

Two single-gender groups, five boys and five girls of the target age groups, from a British school (different from that in the user tests) were involved.

RESULTS AND DISCUSSION

Our empirical data from the survey and user tests reveal some intriguing gender-differences in terms of gameplay habits and gametype preferences, which largely corroborate the findings of previous research studies documented in the literature (e.g., [3]). For the sake of brevity, here we present findings that are relatively distinct from those of related work and unique to 80Days.

Study 1: Gender Differences in Identification with the Main Player Character

With the aim of evaluating to what extent the respondents tended to associate the Boy’s (the main player character) attributes with their own, they were asked to rate first the Boy and then themselves, using a 7-point scale, with respect to six pairs of contrasting adjectives adapted from the instrument Speech Evaluation Instrument [4] consisting of three subscales – *superiority* (Intelligent vs. Unintelligent; Uneducated vs. Educated), *attractiveness* (Friendly vs. Unfriendly; Cold vs. Warm) and *dynamism* (Peaceable vs. Aggressive; Talkative vs. Shy).

The exercises resulted in a set of “Boy-based ratings” and another set of “Me-based ratings”. We computed the correlations among them independently for the German and British samples. A number of statistically significant correlations are found (Fig. 1 and Fig. 2). Results consistently show that the German female respondents tended to perceive the attribute interrelations, be they applied to the Boy or themselves, in a more complicated manner than did their male counterparts. In contrast, the British respondents’ perceptions, irrespective of gender, are less complicated than those of their German counterparts. The British male

respondents tend to perceive the associations in a more complex way than their female ones – a reverse of the trend demonstrated by the German sample.

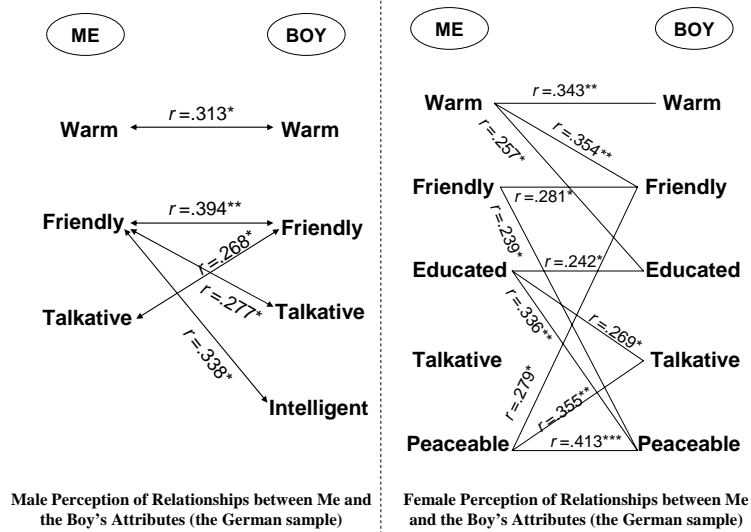


Fig. 1. Gender-specific perceptions of the play character and oneself (German sample)

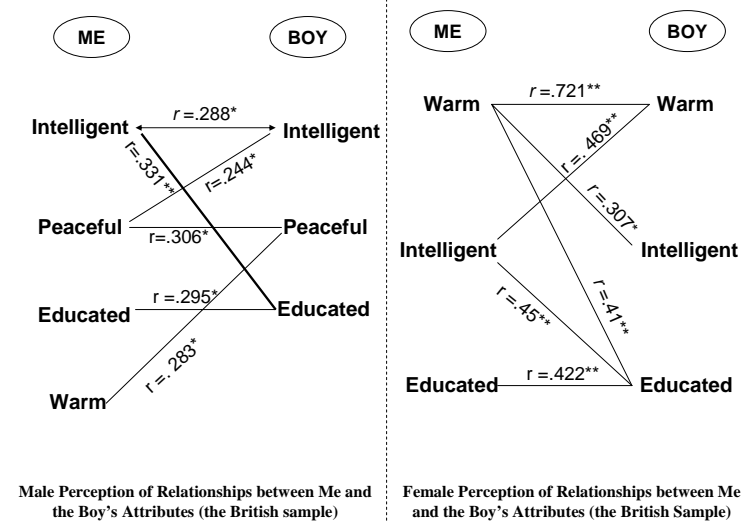


Fig. 2. Gender-specific perceptions of the play character and oneself (British sample)

Study 2: User Tests - Learning Efficacy

Our basic assumption was that by completing the four missions of the game the participating children could gain better understanding of the geographical concepts addressed therein. The improvement was measured in terms of the significant difference in their performance between Pre-ALQ and Post-ALQ. *On average, the children in England and Austria demonstrated statistically highly significant learning gain* (Fig. 3). When breaking down the data by gender, some interesting observations were obtained. For Pre-ALQ, the British and Austria boys performed significantly better than their female counterparts. For Post-ALQ, only the British boys remained significantly better than the girls whereas there was no significant difference between the Austrian boys and girls. Furthermore, the British boys did *not* improve to a significantly larger extent than did the girls. In other words, both genders benefited from the gameplay in terms of knowledge gain, but it did not privilege the boys. In contrast, the Austria girls improved to a somewhat larger extent than did the boys. It suggests that the Austrian girls

could benefit more from the gameplay than did their counterparts. Like the British sample, the game did not privilege frequent gamers.

	British Sample N = 36 (Girls: 16; Boys: 20)	Austrian Sample N = 35 (Girls: 21; Boys: 14)
Sig. diff. in overall learning gain (Post-ALQ > Pre-ALQ)	✓ Gain of 6.6 points ($t=5.3, p<.001$)	✓ Gain of 6.9 points ($t = 5.4, p<.001$)
Gender: Sig. diff. in Pre-ALQ	✓ Boys > Girls ($p<.05$)	✓ Boys > Girls ($p<.05$)
Gender: Sig. diff. in Post-ALQ	✓ Boys > Girls ($p<.05$)	✗
Gender: Sig. diff. in learning gain	✗	✓ Girls (gain of 8.8 points) > Boys (gain of 4.8 points) ($p<.1$)

Fig. 3: Differences in assessment of learning questionnaire by sample and by gender

Study 2: User Tests – Usability and User Experience

Generally speaking, Demonstrator 1 was proved usable. The participants showed high acceptance towards the game, especially the character Feon the alien. They found four micro-missions easy and enjoyable, though they preferred them to be somewhat shorter. The perceived cognitive load induced by the gameplay was generally low.

Evaluation of children’s game experience is addressed along seven dimensions: *challenge, competence, flow, immersion, negative affect, positive affect* and *tension* [5]. Accordingly, a set of statements have been adapted; the participants were asked to rate each of them with a 5-point Likert scale, with the rightmost and leftmost anchors being ‘not true at all’ and ‘very true’.

For the British sample, there were significant differences in two dimensions:

- *Competence*: boys > girls (i.e. the girls rated themselves lower in competence for playing the game);
- *Flow*: boys > girls (i.e. the boys gained stronger positive experience from playing the game);

Similarly, for the Austrian sample, there were also significant differences in two dimensions with one of them being common with that of the British sample:

- *Flow*: boys > girls;
- *Challenge*: girls > boys (i.e. the girls found it more challenging to play the games);

Study 3: Focus Groups – Gender Differences

Discussions in the focus groups were transcribed. Some gender-specific findings were obtained.

- *Game recall*: The girls tended to name more negative features than the boys (i.e. the boys had a higher acceptance towards the game or a weaker memory of the game)
- *Guided discussions* on the most positive and most negative aspects of the game:
 - Most positive: flying the UFO (boys) and colourful graphical presentation (girls)
 - Most negative: not enough actions (boys) and passive listening (girls)
- *Role adoption*: The boys had no problem in recognizing that they were supposed to assume the role of the main character whereas some girls did.

Apparently, gender is one of the critical factors influencing the acceptance and the impact of the game.

CONCLUDING REMARKS

The Demonstrator 1 was a successful deliverable of 80Days, given the following facts:

- the knowledge gain of the children as evident by the significant differences in the learning performance before and after playing the game; some children expressed their appreciation about the education value of the game;
- the usability of the game was generally perceived as satisfactory; the major game components are easy to operate; the level of cognitive load or frustration was generally perceived as low;
- the user experience of the children were generally positive; most of them intended to play the next episode of the game;
- the game character Feon was successfully created;

Besides, the Evaluation Team have acquired some useful practical experiences that will be valuable for the future validation studies. Some of the above findings have been presented and published in prestigious international conferences ([6], [7]). In particular, the gender differences identified invite further research efforts to deepen our understanding of the related issues. Other challenges include:

- Refine evaluation methodologies for the adaptivity features; one particular challenge is to validate the underlying mathematical model with empirical data;
- Develop experimental design for studying visual attention with psycho-physiological measures and their relationships with other factors;
- Explore theoretical models for learning effects of 2D vs. 3D, augmented TAM, and Activity Theory on game-based interaction;
- Identify measurement and structural models for user experience metrics for digital educational games – an emerging research topic;

REFERENCES

- [1] Desurive, H. Caplan, M., Toth, J.A., 2004. Using heuristics to evaluate the playability of games. In *Extended Abstracts of CHI 2004*, 1509–1512
- [2] Korhonen, H. & Koivisto, E.M.I. (2006). Playability heuristics for mobile games, In *Proceedings of MobileHCI'06* (pp. 9-16).
- [3] Kinzie, M.B., & Joseph, D.R.D. (2008). Gender differences in game activity preferences of middle school children: Implications for educational game design. *Educational Technology Research & Development*, 56, 643-663.
- [4] Zahn, C.J., & Hopper, R. (1985). Measuring language attitudes: The speech evaluation instrument. *Journal of Language and Social Psychology*, 4, 113-122.
- [5] IJsselsteijn, W.A., de Kort, Y.A.W., & Poels, K. (in preparation). Development of the Game Experience Questionnaire (GEQ)
- [6] Law, E. L-C., Gamble, T., & Schwarz, D. (2009). Gender and Cultural Differences in Perceiving Game Characters of Digital Educational Games. In *Proceedings of INTERACT 2009* (pp. 149-153).
- [7] Law, E. L-C., Gamble, T., Schwarz, D., Kickmeier-Rust, M., Holzinger, A. (2009). A Mixed-Method Approach on Digital Educational Games for K12: Gender, Attitudes and Performance. In *Proceedings of 5th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society (USAB 2009)*, Linz, Austria, November 9-10, 2009 (pp. 42-54). LNCS 5889