

# Combining Fast-Paced Usability and Scientific Testing to Improve the Lunar Quest Physics Game

Holly Blasko-Drabik, James Bohnsack, and Clint Bowers

University of Central Florida, USA

HollyBlasko@gmail.com, jimisack@gmail.com, bowers@mail.ucf.edu

**Abstract.** This study focuses on conducting fast-paced in-house testing in combination with user comments and scores on the Questionnaire for User Interface Satisfaction (QUIS) to provide a better educational game. The standard QUIS was shortened and focused to be more sensitive to aspects of our educational game. The game, Lunar Quest, was created at the University of Central Florida as a supplemental instruction tool that would provide students with examples of several different Physics concepts while being an enjoyable and fun game. Overall the modified version of the QUIS was not successful in determining which categories of the game should be targeted, although the open ended questions did help our researchers focus the game redesign and showed improvement throughout the testing period.

**Keywords:** Usability, QUIS, Serious Games, Learning.

## 1 Introduction

We have all heard that usability and in-house testing can improve a product, making it more user friendly or effective. The counter argument is that such testing often takes too much time, effort, and money to make it worthwhile. This study focused on conducting fast-paced in-house testing in combination with user comments and scores on the Questionnaire for User Interface Satisfaction (QUIS) to provide a better educational game. The game, Lunar Quest, was created at the University of Central Florida as a supplemental instruction tool that would provide students with examples of several different Physics concepts while being an enjoyable and fun game. Lunar Quest currently consists of 8 mini-games, each pair of games teaching one main concept typically covered in general Physics courses, such as the effects of gravity or kinematics.

The QUIS was specifically designed to determine the user's satisfaction with computer software in 1988 as a nine point Likert scale questionnaire. Users are asked to evaluate each question on this scale based on a pair of adjectives at the anchor points. For example, "Error messages were unhelpful (1) to helpful (9)", if the users felt that the question did not apply to the system then they are asked to circle N/A. Since it was first implemented the measure has been increased to cover many new aspects of technology. The current version of the QUIS is broken down into 12 parts and contains 125 Likert scale question, 12 open ended response questions, and several questions pertaining to the users past experience with the system

or similar systems. Because we wanted to administer the QUIIS after each of the mini-games we were looking for a shorter more focused version of the QUIIS that could be completed with 5-10 minutes but that would still help us focus our development of the mini-games.

In order to achieve this goal, a modified version of the QUIIS (see appendix a), specifically targeting five different categories we believe would impact the learning and usability of the game was created and given to our game testers. These categories were based on the full 12 part QUIIS many subsections and questions were removed. It was determined that 1) Overall reaction to the game, 2) Text and Organization, 3) Use of terms and information, 4) Learning, 5) Game features & Capabilities would be targeted. Many questions within these categories were also removed if they seemed confusing to the users or experimenters or did not pertain to features that were included in the games we were testing. The final version of the questionnaire that we used consisted of 28 of the Likert questions. Researchers also asked users to several open ended questions, specifically, what they liked or disliked about the game, how we could make the game more usable, and what would make the game more fun or interesting.

## **2 Method**

Pairs of students were brought into the lab and asked to play the mini-games and evaluate them using the modified QUIIS. Ideally, between 10 and 14 students were included in each phase of testing, however, due to the discovery of several major bugs or fixes that took longer than expected to implement, the groups were not equal. Testing was broken up into 5 phases (including the initial pilot testing) and 64 students played the Lunar Quest mini games. In the first period of testing, critical bugs and errors were discovered, recorded, summarized, and provided directly to the game developers. As soon as these changes were made, another small group of subjects were tested and the games were again modified. This allowed the mini-games to be quickly improved while development was ongoing instead of waiting until the games were completed to test them.

## **3 Results**

One-way ANOVAs were conducted looking at the five categories across all eight mini-games. The ANOVAs showed that there was a significant improvement in several of the mini-games, specifically the game features and the way information was organized. Follow up t-tests also showed several significant findings where a decrease in scores between the phases of the game were found, specifically on the learning and capabilities categories.

## **4 Discussion**

Overall, many important issues were discovered early enough in the design process to allow for the games to be improved without a total redesign. By using this method, we

were able to identify which specific categories our users found most important, specifically the game features the way information was organized.

Using the modified QUIS allowed us to target suspected features that we believed to be critical in learning the material and concepts explained in the games. However, based on our results we do not believe that the modified QUIS was sensitive enough to accurately determine changes in the usability of the game. The student's comments to the open-ended questions showed a marked improvement in their opinions of the games between many of the phases of testing, however, the data showed that no significant change or an inverse relationship.

Several possible constraints may also have influenced our results. In our efforts to shorten the questionnaire to 5 or 10 minutes, several questions that many have been important to the validity of the overall QUIS may have been removed. Because the game needed to be completed and changed quickly, group sizes varied from 5 to 15, in addition several students data could not be used because they circled the adjectives instead of a number on the 0-9 scale.

Future testing of the Lunar Quest game is currently being conducted looking not only at the usability of the games, but at whether using the game has a positive impact on learning the targeted Physics concepts.

## Appendix A

Overall reaction to the game														
1		terrible	0	1	2	3	4	5	6	7	8	9	wonderful	N/A
2		difficult	0	1	2	3	4	5	6	7	8	9	easy	N/A
3		frustrating	0	1	2	3	4	5	6	7	8	9	satisfying	N/A
4		dull	0	1	2	3	4	5	6	7	8	9	stimulating	N/A
5		rigid	0	1	2	3	4	5	6	7	8	9	flexible	N/A
Screen														
6	Reading characters on the screen	hard	0	1	2	3	4	5	6	7	8	9	easy	N/A
7	Highlighting simplifies task	not at all	0	1	2	3	4	5	6	7	8	9	very much	N/A
8	Use of bolding	Unhelpful	0	1	2	3	4	5	6	7	8	9	Helpful	N/A
9	Organization of information	confusing	0	1	2	3	4	5	6	7	8	9	very clear	N/A
10	Arrangement of information on screen	Illogical	0	1	2	3	4	5	6	7	8	9	Logical	N/A
11	Sequence of levels	confusing	0	1	2	3	4	5	6	7	8	9	very clear	N/A

Terminology & Game Information														
12	Use of terms throughout game	inconsistent	0	1	2	3	4	5	6	7	8	9	consistent	N/A
13	Terminology related to task	never	0	1	2	3	4	5	6	7	8	9	always	N/A
14	Position of messages on screen	inconsistent	0	1	2	3	4	5	6	7	8	9	consistent	N/A
15	Prompts for input	confusing	0	1	2	3	4	5	6	7	8	9	clear	N/A
16	Computer informs about its progress	never	0	1	2	3	4	5	6	7	8	9	always	N/A
17	Error messages	unhelpful	0	1	2	3	4	5	6	7	8	9	helpful	N/A
Learning														
18	Learning to play the game	difficult	0	1	2	3	4	5	6	7	8	9	easy	N/A
19	Exploring new features by trial and error	difficult	0	1	2	3	4	5	6	7	8	9	easy	N/A
20	Remembering directions	difficult	0	1	2	3	4	5	6	7	8	9	easy	N/A
21	Performing tasks is straightforward	never	0	1	2	3	4	5	6	7	8	9	always	N/A
22	Help messages	unhelpful	0	1	2	3	4	5	6	7	8	9	helpful	N/A
23	Instructions	confusing	0	1	2	3	4	5	6	7	8	9	clear	N/A
Game Capabilities														
24	Game speed	too slow	0	1	2	3	4	5	6	7	8	9	fast enough	N/A
25	System reliability	unreliable	0	1	2	3	4	5	6	7	8	9	reliable	N/A
26	System tends to be	noisy	0	1	2	3	4	5	6	7	8	9	quiet	N/A
27	Correcting your mistakes	difficult	0	1	2	3	4	5	6	7	8	9	easy	N/A
28	Designed for all levels of users	never	0	1	2	3	4	5	6	7	8	9	always	N/A