Games as Skins for Online Tests

Srinivasan Ramani, Venkatagiri Sirigiri, Nila Lohita Panigrahi, Shikha Sabharwal
HP Laboratories
HPL-2008-92

Keyword(s):
multiplayer and social game development, collaboration competition and community, foundation and theory for design

Abstract:
Games play a dual role: they test the player's competence, and at the same time provide learning opportunities. They offer a simple form of a reward – the pleasure of playing, without serious consequences for failure. Tests can be converted into games. In this context, it is valuable to create rich "skins" in which tests can be clothed with no major effort. This paper reports research aimed at improving the scope of skins. We have primarily restricted ourselves to skins that accept a test consisting of objective type questions and use it to provide students with a gaming environment. A number of issues are addressed in the context of designing such a system, including collaboration, content creation, content grading, reward structure, team building and personalization. After a discussion of these issues, we proceed to describe features we were able to incorporate in a specific design inspired by the cricket metaphor, which we have implemented. We have started using this game for educational testing at a university level program in engineering.

External Posting Date: July 21, 2008 [Fulltext] Approved for External Publication

Internal Posting Date: July 21, 2008 [Fulltext]

Submitted to Digital Game and Intelligent Toy Enhanced Learning, November 17-19, 2008 Banff, Canada

Games as Skins for Online Tests

Srinivasan Ramani,
International Institute of Information Technology and HP Labs India, Bangalore
Ramani@iiitb.ac.in

Venkatagiri Sirigiri,
Nila Lohita Panigrahi, Shikha Sabharwal
HP Labs India

Abstract

Games play a dual role: they test the player's competence, and at the same time provide learning opportunities. They offer a simple form of a reward – the pleasure of playing, without serious consequences for failure. Tests can be converted into games. In this context, it is valuable to create rich "skins" in which tests can be clothed with no major effort. This paper reports research aimed at improving the scope of skins. We have primarily restricted ourselves to skins that accept a test consisting of objective type questions and use it to provide students with a gaming environment. A number of issues are addressed in the context of designing such a system, including collaboration, content creation, content grading, reward structure, team building and personalization. After a discussion of these issues, we proceed to describe features we were able to incorporate in a specific design inspired by the cricket metaphor, which we have implemented. We have started using this game for educational testing at a university level program in engineering.

1 Introduction

The value of educational games is widely recognized [1] [2]. In this note we will deal with a form educational games we name "Guiz", using a word synthesized from "game" and "quiz". Here quizzes are made into games giving "low value" rewards to encourage learning. The reward would be a token rather than anything substantial such as a college scholarship. The idea is to "de-risk" playing and to take the fear of losing, out of the game. We explore possible and desirable objectives for such games that are practical with the current state of the art in technology.

Our system implementing many of the features described here is running as a Beta on the learning management system Moodle [3]. Students have individual login accounts and can take tests, including those thinly disguised as games.

2 The Use of "Skins"

Most teachers would not like to spend a lot of time creating educational games, but might be willing to use them if they can be "customized" for the context in which it is going to be used. In this paper, we assume that many of them would be willing to use software which takes a test in a standard form and enables students to play a game based on the test. We think of the game as a skin which the teacher can put on a test to make it attractive. *The development of such skins is very important if educational games are to be used in serious educational programs.*

A widely used type of tests is the Objective Form Test (OFT). We name an item of software an OFT Skin (OFTS) if it takes an OFT and converts it into a game. We explore features of OFT skins that would offer a variety of advantages, and go on to describe a skin we have designed and implemented. Design and development of rich "skins" incorporating several techniques of learning value is feasible because good skins are likely to be reused with a large set of tests on a large number of campuses. We believe that developing rich skins with a focus on educational tests could be a useful first step. Developing more comprehensive skins which can support a variety of learning activities could be made possible by lessons learnt during the first step. Integration of the instructional and testing objectives and use of an impressive visual environment heightening the competitive spirit cannot be ignored in games; but they could be considered to be of lower priority in the early stages of skin design. Considerable design and development effort is therefore justified. In fact, a skin designer needs to consider, prior to the design effort, all possible features that could be consistent with his game-model. There are a number of skins in use, for instance, the Snakes and Ladders game [4] distributed

by the Moodle Community [5]. However, there is plenty of room for innovation in skin design.

3 Desirable Features of Skins

All skins cannot provide all features. However, integrating a number of the following features into a skin makes it valuable to instructors using them. Our focus will be on skins for tests as we discuss these features.

3.1 Learners and Teams

Designing on-line educational game skins for supporting a large number of players at any given time has significant advantages. It maximizes the returns when these skins are used on one or more large networks. It increases the probability that someone would be playing at a time a learner wishes to join in. A corollary is that it is useful to design a game for playing over the Internet, ideally using nothing more than a browser. While some institutions may limit access to their game server only to those on the LAN, others may permit access over the Internet.

It is also desirable to support different tests to run concurrently on the same server, utilizing a common skin. This will avoid regimentation and allow the learners a choice of games.

A game such as cricket has about half the players sitting in the pavilion at any time! Only two players from the batting side are active. Most of the fielders have little to do most of the time! It is best to avoid this and keep as many players as possible in a highly active state.

Team play is not always possible. Sometimes a student may need to use an educational game alone. One way to do this without altering the structure of the skin is to make virtual players available whenever needed. Virtual players could perform at pre-set levels of competence; these could be defined by specifying means and standard deviations for the scores that they should get. These statistical expectations can be ensured by appropriately controlling pseudo-random processes involved in the implementation of the virtual player facility.

Students could be allowed to choose the virtual players they wish to play. These virtual players could have the names of popular champions to add color to the game!

3.2 Competition

A team structure in a competitive game encourages collaboration and peer learning. Since team members wish to maximize the team's performance, there is an incentive for extending support to team-mates. Under what circumstances are relatively homogeneous teams better? Where do non-homogeneous teams perform better? Should learners be allowed to form teams freely, or should we assign learners to teams systematically/or at random? These issues require further research.

Competition is a major factor introducing *affect* into what would otherwise be purely cognitive activity. The role of *affect* is central to the attraction of games. Hence, it is desirable to promote some degree of competition between individuals or between teams, whenever possible.

An interesting form of competition is the tournament in which competitors play against each other pair wise, which is possible in a number of games. Each such "match" defines a useful ordering that can be used as a tool to compare the relative merit of teams. In turn, earning a good rank becomes the motivator to play. Using a sequence of matches, we can rank contestants. One way of doing this is as follows:

a) Assign each player a score of 1 to start with
b) Pair the players for the purpose of the next match, pairing players with similar scores as far as possible. If more than 2 players have roughly equal scores, pair them randomly among themselves. Arrange for all pairs to play simultaneously. Two players within a pair will play against each other.
c) Increase the score of the winner in each pair by 1, leaving the loser's score unchanged.
d) Repeat the sequence of steps (b) – (c) some K times; this implies that every participant will play K matches. This means that each player will take K different tests.
e) The scores after K repetitions define a partial ordering of the contestants.

An important feature of this design is that a poor player is not eliminated from the tournament after one match. Every one will play N matches.

3.3 Feedback

Players get very quick feedback in most games. When they make a mistake or win a few points, they know this immediately. Such immediate "knowledge of results" enables the student to learn as a result of answering a question. This will also allow the student to move on to the next question that may have a different focus as compared to the earlier one, without

having to keep notes to return to later. Skins should offer similar real-time feedback through quick evaluation of student responses. This is easy for objective type questions. In other cases, when the question is not of objective type, the system might have other participants evaluate a student's response in a game-like situation. Such evaluation would have to be done in real time.

The system could reveal the correct answer to the student if he asks for it after he learns that he has given a wrong answer. The system could also add a short explanation as to how the preferred answer is the right answer.

Giving early feedback to the designers of new questions (these may be the learners themselves; see Sec 3.4) is an interesting challenge, which can be met by real-time "item analysis". We can assume that a score S_i is available for every player in a game at every stage during its course, at least the default score of 0 for every player at the start of the game. Alternatively, we could use a cumulative score, such as batting average, from previous games. Consider a question Q_j immediately after it is answered by the players. Let P_R be the mean of scores of all those who successfully answered Q_j and P_W the mean of scores of all those who did not. We find that $(P_R - P_W) / (P_R + P_W)$ is an easy-to-compute value related to the question answered and that it grades "good questions" above the "poor questions". We can use this Figure of Merit (FoM) associated with a new question to give feedback to the designer of questions, even immediately after the first-time use of his question. This, however, requires that the number of players in the game is adequately large to provide a statistically significant test population.

3.4 Collaboration and Content Creation

Collaborative learning is a widely researched topic [6], with many facets. It is valuable for skins to motivate learners to offer their teammates questions, hints and explanations. This motivation can also encourage learners to contribute content useful to a game skin. Our focus here is on a special form of collaboration in which test performance of every team member is a matter of concern for the whole team. We expect that in this situation, collaboration would occur well in advance of a test thinly disguised as a game. We expect that team members would freely help each other prepare for the game, offering consultation and tutorial help as needed.

The system could display a 'canned' hint when the learner submits a wrong answer, if there is enough time left in the time slot allotted for that question. This hint could simplify the task of the learner by giving additional information, or a simple explanation of why the answer he gave is wrong. A hint would enable him to make a second attempt at the question, for earning at least reduced marks for that question.

This hint need not necessarily come from the question designer, or be stored along with the questions. It could be entered by the team-mates who answer the questions early in the time slot and correctly. These answers could be stored for display to the players who need it, only after they enter their own answers. Alternatively, the skin could pass on to a learner hints from his team-mates late in the time slot, that is after it is clear that he is facing difficulties in answering the question.

Explicit sharing of answers, for instance by typing "c" as a hint for a multiple-choice question, can be discouraged through an honor system and a penalty. Knowledge that hints are being stored for later examination by invigilators could be sufficient to discourage dishonest practices.

Hints given by students and stored in this manner can also be transferred, after scrutiny, to an FAQ. The hint FAQ could be external to the question database, and merely consist of hints linked to question IDs. This FAQ can provide the "canned hints" referred to earlier.

Learners can go beyond creating hints. They could be incentivized to create questions through an appropriate game structure. These questions would behave like challenges in a game, for instance the kind of challenge a pitcher in a simulated baseball game poses to the hitter. Motivating students to create questions has been recognized to be valuable [7], [8]. Such student created questions can be evaluated statistically on the basis of student responses. Good questions identified could be stored in a central database after the game for future use. The skin could incorporate some statistical tools. Alternatively, a skin could use a separate "item analysis" module to evaluate questions.

It is worth noting the special value of learners being involved in the creation of hints and questions. This goes well beyond the quantitative increase in the number of content creators. Such constructive activities support the learning process and critical thinking. They also give the learner the incentive of being the proud creator of something of value to the others, and of course opportunity to show off a little!

3.5 Synchronization and Personalization

The focus of work reported here is on games based on objective type tests. To ensure that the

competitive structure of the game is preserved, it appears desirable that all participants should have the same question delivered to them simultaneously. It is also desirable that they all have to answer the question within the allotted time. This requires that adequate time has to be given to every participant. A suitable measure for this is T_{90}, the time decided by prior testing of the question and found to be adequate for practical purposes. The timing test could give a reasonably long time for a sample population to answer the question under test and compute T_{90} as the time within which those eventually answer the question correctly submit their answers. While pre-testing is useful in providing initial estimates of T_{90} it is possible to improve these estimates by using data from repeated use of the questions.

A related problem is to avoid boredom for those students who submit answers quite early, and have to wait for the time slot to be over. A possible solution to this problem is to let them use the spare time to offer help to their team-mates as discussed in Section 3.4.

Does synchronization necessitate that the same question should be presented to every player in a given time slot? Or, can we present different questions to different players during the same time slot? One can argue that even if different questions are presented to different players, the time slots in which they need to reply should be the same, requiring the system to use estimates of T_{90} to select a set of questions for each time slot ensuring that they require roughly equal time to answer. In the interest of fairness, the set of questions presented during a given slot should also be at roughly equal levels of difficulty.

The possibility of presenting different questions to individual learners (or members of individual teams) opens up the possibility of personalization. The system could create and regularly update student models on the basis of questions learners have answered or failed to answer. This could be used to identify required areas in which a learner's competence is to be tested, at a given point in time. This could enable the system to select an appropriate question to be asked during the next step.

3.6 Impressive Visuals

An enriched visual and auditory environment could be valuable in some cases, making the game more enjoyable and realistic. Learning about anything with a visual dimension such as dynamics would be made easier. A technical objective would be to create a well-integrated visual and auditory environment that goes beyond the use of individual video-clips.

We should note another possibility. An impressive visual environment might be valuable in games involving sensory-motor skills and in entertainment oriented games. It is possible that such an environment would distract the learner from the essence of a Guiz, which is more the quiz than a game!

3.7 Encouraging Poor Performers

Guiz's should ideally be designed to retain the interest of learners at all levels of achievement [9]. Declaring high performers alone as "winners" every time and giving them all the rewards could result in the "losers" withdrawing from the competition to avoid disappointment. The phrase "small carrot, big stick" sums up the possible negative effect of formal tests on learners with lower scores.

4 Guiz C1: Cricket-like OFTS

We have designed a skin, named Guiz C1, which uses concepts familiar to cricket players, but deviates significantly from the rules of cricket. Why cricket? One reason was that bowling and batting are complementary activities that can be mapped on to asking questions and answering them. Other reasons include its team structure, as well as the game's international popularity.

Guiz C1 is a multi-player competitive game and serves as a test bed for trying out a number of techniques discussed in Section 3. We map asking questions onto "bowling" and answering them onto "batting". The game utilizes a bank of questions stored in Moodle.

Our implementation has a built-in mechanism to estimate the level of difficulty of any new question during its first use, in terms of the percentage of players who give the correct answer. It also has a mechanism to compute the Figure of Merit mentioned earlier in Section 3.3. These mechanisms are used to give appropriate scores to bowlers and batsmen.

4.1 Competition

Currently a competitive situation is created for each player by giving "runs" to the batsman and a bowling score to the bowler. Averages are maintained and displayed for batting as well as bowling. However, there is no notion of the bowler "taking a wicket", thereby sending the batsman out of the game. As mentioned earlier in Section 3.2, tournaments in which all teams can compete for comparative ranking is possible. However, our system does not at present

provide for any automation of the required data handling for such tournaments.

Virtual players are provided for, and they use the questions from the question bank when they bowl. The scores of a virtual player during batting are determined by pseudo-random number generators with specified mean and standard deviations. Specified bowling scores for virtual players makes them 'difficult bowlers' and 'easy ones'. The difficult ones choose questions at a higher level of difficulty.

4.2 Team Structure

Teams of from two to four members and individual players are permitted. However, a large number of teams are permitted, thereby supporting a large number of players. Currently our code limits this to 500, but the design permits easy extension to a larger number of players. The limits are likely to be set by the hardware available, rather than by the design. The system also allows the administrator to schedule/initiate multiple games each of which has its own set of players.

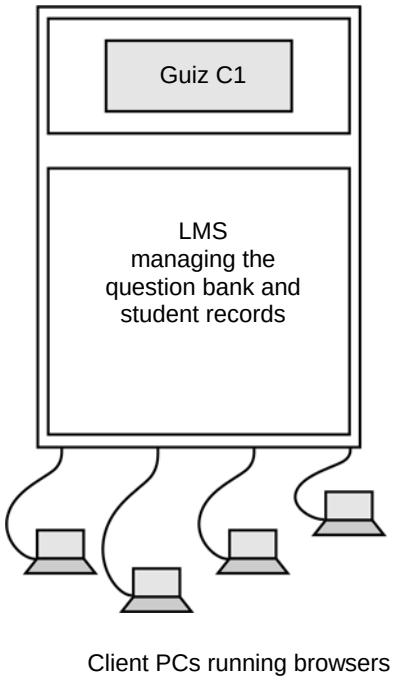

Figure 1. System Overview of Guiz C1

4.3 Timely Feedback:

A learner answering a question correctly gets an immediate reward in the form of his increased "batting average". This is possible because the question bank associates an expected answer and marks to be awarded for answering it with every question. However, consider the student who has designed and stored a set of new questions in the question bank for his use. It is not satisfactory to let him decide marks to be given for each of his questions using his subjective judgment. Some degree of testing the questions would be necessary before it can be given an appropriate "weight". We describe, in the next sub-section a method of assigning what we consider appropriate initial marks to a question "on the fly". These marks can be refined as the question is used again and again, providing us with greater statistical data. The marks associated with a question should not simply depend on its level of difficulty. In our design it depends on the figure of merit defined in Section 3.3.

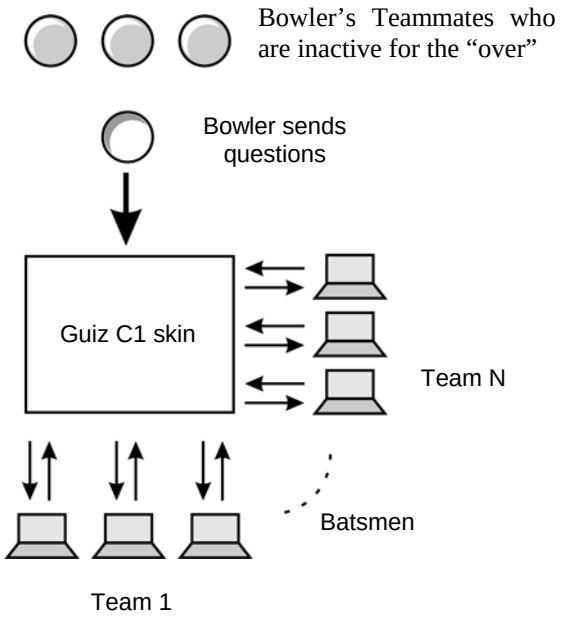

Figure 2. Players in Guiz C1

4.4 Learners as Contributors

At any given moment of play, one player is the designated bowler. Members of the bowler's team are not allowed to bat, but every other player can bat.

The bowler is identified well in advance of the game by someone serving as the administrator. The bowler, therefore, has time to select questions from the question bank for his use during bowling. Alternatively, he can enter into Moodle the new

questions he has designed in advance of the play. He submits a list of question numbers to be used in bowling during his turn, and the system does everything else automatically.

The bowler does not even have to be logged in when his balls are bowled. Giving an incentive to learners in the form of increases in bowling scores motivates them to be creators of useful content in the form of contributions to the question bank.

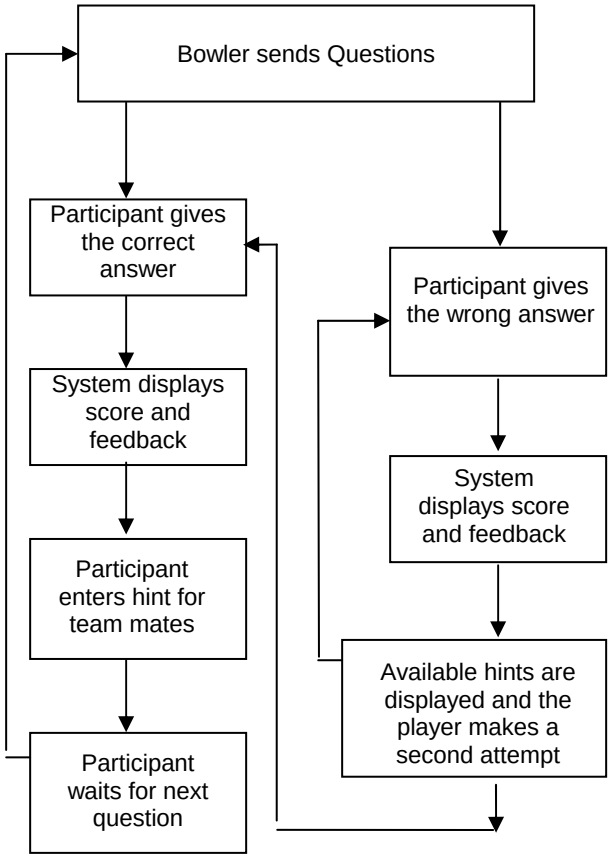

Figure 3. Possible sequences of events

4.5 Use in a University Level Course

Our first use of Guiz C1 is with a short self-study course for incoming graduate students at the International Institute of Information Technology, Bangalore. The students are expected to spend a number of weeks taking three "preparatory courses" to brush up their undergraduate level knowledge. One of these is a web-based course "System Analysis and Design" by Prof. V. Rajaraman, created under the National Project on Technology Enhanced Learning – NPTEL [10]. We took the multiple choice questions covering 11 modules in this course and loaded them for use with Guiz C1. This course is in progress. Table 1 lists the features provided by the system in use.

Table 1. Features Implemented at Present

S. No	Features	Implementation Status
1	Competition between teams	Implemented
2	Allowing large number of players	Implemented
3	Allowing multiple games to run simultaneously	Implemented
4	Virtual players	Implemented
5	Tournaments	No
6	Feedback to individual and to teams	Implemented
7	Canned hints	Implemented
8	Hints from team members	No
9	Learners as content contributors	Implemented
10	Permitting second attempts	Implemented
11	Collecting statistical data on questions	Implemented
12	Synchronization	Implemented
13	Personalization	No
14	Visuals	Stills and video clips
15	Encouraging poor performers to continue playing	Implemented
16	Form of question accepted	Multiple Choice Questions only

5 Conclusion

We have discussed the concept of game-skins for making objective tests more interesting to learners. We have argued that a variety of features have to be built into such skins to increase their utility for instructions. We have described a specific implementation and use of a cricket-like game-skin. We hope to release this skin in the open source form and to offer demos over the Internet.

6 Acknowledgements

Kremena Diatchka was a member of our team when this work was in the planning stages. She did a very useful literature survey and contributed to our discussion meetings.

We acknowledge with thanks the full support given to this project by HP Labs India. The Director, Ajay Gupta gave us access to the computing

environment in which this work was carried out. Even more valuable has been his continued interest in technology in education. The first, third and fourth authors thank the International Institute of Information Technology, Bangalore, and its Director S. Sadagopan, for the support extended to their work in this field.

This research is funded by HP Labs India.

7 References

[1] Kurt Squire, Henry Jenkins, "Harnessing the Power of Games in Education", InSight, Vol. 3, VISION 5, 2003

[2] Sanford Richard, Williamson Ben, "Games and Learning", A Handbook from FutureLab, 2005 http://www.futurelab.org.uk/resources/documents/handbooks/games_and_learning2.pdf

[3] William H Rice IV, "Moodle: E-Learning Course Development", Packt Publishing, 2006

[4] Documentation on Game module http://docs.moodle.org/en/Game_module

[5] The Moodle Home Page http://moodle.org

[6] Jan-Willem Strijbos, Paul A. Kirschner, Rob L. Martens (Eds.), "What We Know About CSCL and Implementing it in Higher Education", Series: Computer-Supported Collaborative Learning Series, Vol. 3, 2004, PP. 280.

[7] Fu-Yun Yu, Yu-Hsin Liu and Tak-Wai Chan, "A web-based learning system for question-posing and peer assessment", Innovations in Education and Teaching International Vol. 42, No.4, November 2005, pp. 337-348

[8] Rosenshine Barak, Meister Carla, and Chapman Saul, "Teaching Students to Generate Questions: A Review of the Intervention Studies", Review of Educational Research 1996, Vol. 66, No.2, PP. 181-221

[9]Winston M.C. Wu, Hercy N.H. Cheng, Mong-Chen Chiang, Yi-Chan Deng, Chih-Yueh Chou, Chin-Chung Tsai, and Tak-Wai Chan, "AnswerMatching: A Competitive Learning Game with Uneven Chance Tactic", The First IEEE International Workshop on Digital Game and Intelligent Toy Enhanced Learning (DIGITEL'07), 2007

[10] Course on "Systems Analysis and Design" by Prof V. Rajaraman of Indian Institute of Science, Bangalore. at http://nptel.iitm.ac.in