Designing and remotely testing mobile diabetes video games

Jonathan DeShazo*, Lynne Harris†‡, Anne Turner†‡ and Wanda Pratt‡§

*Department of Health Administration, Virginia Commonwealth University, Richmond, Virginia; †Department of Health Services, University of Washington, Seattle, Washington; ‡Division of Biomedical and Health Informatics, University of Washington, Seattle, Washington; §Information School, University of Washington, Seattle, Washington, USA

Summary
We have investigated game design and usability for three mobile phone video games designed to deliver diabetes education. The games were refined using focus groups. Six people with diabetes participated in the first focus group and five in the second. Following the focus groups, we incorporated the new findings into the game design, and then conducted a field test to evaluate the games in the context in which they would actually be used. Data were collected remotely about game usage by eight people with diabetes. The testers averaged 45 seconds per question and answered an average of 50 total nutrition questions each. They self-reported playing the game for 10–30 min, which coincided with the measured metrics of the game. Mobile games may represent a promising new way to engage the user and deliver relevant educational content.

Introduction

Diabetes is a major health challenge. We know that diabetes education improves quality of life and clinical outcomes[1–3] and can also reduce the overall financial burden of the disease.[4] However, conventional diabetes education can be expensive[5] and inconvenient for some people.[6] In an effort to increase the popularity and effectiveness of diabetes education, researchers have explored using video games to deliver diabetes education.[7]

Video games are played by almost all teenagers[8] and nearly half of adults.[9] Using video games as an educational delivery method has advantages when compared with conventional methods. First, video games are intrinsically motivating and therefore may be more effective.[10] Furthermore, tailoring education to the individual's characteristics increases its effectiveness,[11] and video games can tailor the message to the user's needs automatically.

Mobile phones can support relatively sophisticated game play with rich content. Furthermore, they have the added advantage of portability. Mobile phones are also very common. In the USA, more than 90% of adults and 79% of teenagers have mobile phones.[12] We have investigated game design and usability for video games to deliver diabetes education. We conducted formative research to tailor the design to the needs of our target community, and usability testing to further refine the user experience.

Accepted 2 March 2010
Correspondence: Jonathan DeShazo, Department of Health Administration, Virginia Commonwealth University, Grant House Room 205, 1008 East Clay Street PO Box 980203, Richmond, Virginia 23298-0203, USA (*Fax: +1 804 828 1894; Email: jpdeshazo@vcu.edu*)

Methods

The educational goals of the mobile games were the same as the feature rich desktop games, i.e. to increase nutritional-estimation and food-comparison skills. We investigated three mini-games with brief and simple game play: Hangman, QuizShow and Countdown (Figure 1).

Our design strategy was to create simple casual games with nutritional goals embedded within the game. We supported this design strategy using elements from health communication theory and education theory including tailoring,[13] scaffolding,[14] interactive learning techniques,[15] and linking educational attainment with score.[16] It is well established that conventional diabetes dietary education is more effective when tailored to the individual.[17] Accordingly, the mini-games can be customized by diet (vegetarian or not), nutritional goals (carbohydrates, calories or energy density), and skill level (easy or hard) to suit the dietary needs, preferences and skill levels of the player. The nutritional information in the games was adapted from the US Department of Agriculture Food and Nutrient Database for Dietary Studies.[13]

Games

In Hangman, nutritional education is framed within the classic guessing game. Players have six guesses to estimate

 DOI: 10.1258/jtt.2010.091012

Figure 1 The mobile video games Hangman, QuizShow and Countdown (left to right) on an iPhone, Windows Mobile and Blackberry, respectively

the calories or carbohydrates in a given food before a stickman is hung by the gallows. QuizShow is based on popular television game show themes. Players choose the correct answer to a nutrition question and have three 'lifelines' available for help during the game. The game play models in Hangman and QuizShow both reinforce nutritional estimation skills while playing a familiar game. However, they are also designed with elements of interactive learning theory embedded within the game play. When a Hangman player guesses the wrong quantity, the game responds with feedback indicating that the guess was either too high or too low. This directional feedback, or 'scaffolding', helps learners progress to a correct answer.[14] Similarly, QuizShow's hint buttons, called 'lifelines' in the game, have been shown in other educational video games to increase participation and learning effectiveness.[15] The third game, Countdown, focuses on food comparison skills. Players choose between two foods based on a given nutritional category (e.g. energy density or calories), and the player receives more points for faster correct responses.

We created initial mockups depicting these game play strategies and interfaces for use in the focus groups. The mockups were designed with actual implementation in mind, which added realistic constraints to the designs. For example, game controls were limited to those commonly available on mobile phones.

Formative research

We used two focus groups to refine the initial design of the games and tailor the interface and game play to the needs of users. We chose focus groups because they are useful in testing new ideas and eliciting a group's attitudes. The focus group participants were recruited from a college campus and a local diabetes clinics.

A moderator facilitated each of the focus groups with the help of an assistant. Both moderator and assistant had previous focus group experience. The assistant actively recorded nonverbal cues, noteworthy concepts, and compared notes and experiences with the moderator. The two focus groups were conducted within a three week period, using the feedback of the first group to partially inform the second focus group. Focus group audio was recorded and professionally transcribed for analysis.

The participants were given an introduction to the objectives of the study, the nutritional goals of all of the video games, and the reinforcing role that the mobile games play. Thereafter, each mini-game was described to the focus group using mockup screens of the games, and discussed in detail.

In part, the aims of the focus group were to elicit the groups' thoughts towards the game's features, game play, scoring and difficulty level. The moderator paid particular attention to eliciting suggestions for improvement and perceived barriers for use along these dimensions. In addition to design topics, the focus groups were also intended to identify factors relevant to the nutritional knowledge and goals of the participants.

Usability testing

We used a mixed-methods approach to remote usability testing by supplementing the automatically collected analytic metrics with a questionnaire to assess the user's attitudes and self-reported experiences. Text communication and multiple choice responses have been found to be relatively effective at capturing critical user events[18] when compared to analytic metrics. The analytic metrics were asynchronously transmitted from the user's mobile device to a server via the mobile device's data connection. Users have the ability to prohibit the game from transmitting the data to avoid consuming data bandwidth or service time. This communication setting is on the device itself and can block communication temporarily or permanently depending on the user's preferences. Game play metrics include the mini-game being played, the current question, the current score, nutrition category, difficulty level and time. From these values, we could measure game play duration and percentage of correct answers. The number of game play metric items were kept low to reduce data transmission volume. Timestamps were established on the server side of transmission. Because the data communication was asynchronous, delivery unconfirmed, and timestamps were made at the server, some transmissions might have been lost and timestamps may have been off by several seconds. We did not anticipate the dramatic lag in mobile device network communication networks compared to broadband Internet and found no indication of this in the literature. This would later be identified as a limitation in our approach.

In addition to the metrics, subject testers completed an 11-item questionnaire with categorical and unstructured (narrative) responses. Categorical questions elicited user's estimates of level of enjoyment, nutritional relevance, amount learned, level of challenge, simplicity of game play and presence of any confusing elements. Open-ended narrative responses were obtained for favourite and least favourite aspects of games, as well as suggestions for improving the games.

All aspects of usability testing were accomplished online. Following approval from the appropriate ethics committee, subjects were recruited for remote testing via advertisements on online message boards specific to diabetes and local classified advertisements. Subjects were required to be aged 18 years or over with a diagnosis of diabetes. Subjects were also required to own a phone that was compatible with the games (i.e. Windows Mobile devices, iPhones and most Blackberry models). Subjects were remunerated with a $10 gift certificate. Subjects downloaded and installed the games themselves. They were instructed to play the games as much or as little as they chose for one week, and then to complete an online questionnaire.

Results

Formative research

Six people participated in the first focus group and five in the second. The participants were adults, with Type 1 ($n = 8$) and Type 2 ($n = 3$) diabetes. The focus group dialogue helped to refine the difficulty level, personalization categories and scoring system of the games. The groups also provided their perspectives of the advantages and disadvantages of mobile games in general, as well as the three mini-game ideas in particular.

Overall, each of the mini-game concepts was well liked by the focus groups. In general, participants could see the potential value of nutritional knowledge reinforcement training on phones, though not all participants felt that they would play the games. One participant felt that video games should be targeted towards young people. Several members of the focus groups thought they would play the games if their phones supported them. However, two individuals were not interested in using their phones for activities other than talking.

The focus groups also shed light on the appropriate length of the games. Players expect a much shorter game play on the mobile phone games than on desktop games. They associate mobile phone games with passing brief periods of idle time during the day. For example, mobile phone games were used by focus group members to pass time while waiting for someone or during a commuter journey. Their emphasis was on brevity and the ability to quickly pick up the game and put it down. One estimate of the maximum duration of a game was five minutes.

The concept of game difficulty was discussed in detail. Participants associated difficulty level with the completeness of the food description, the time the respondent has to answer, and the range of acceptable nutritional scores listed as answers. The best strategy for setting question difficulty was to vary a percentage buffer. For example, in 'hard' difficulty mode in Hangman, a player would have to guess the nutritional value within 5%. Similarly, the multiple choice values in QuizShow would be spaced apart based on 5% of the correct value. These buffer zones must be known to the player to be useful. We relied on conducting alpha testing as well as more comprehensive usability testing (described below) to further refine the difficulty levels that were initially identified by the focus groups.

Remote testing

Following the focus groups, we incorporated the new findings into the game design, thereby completing the initial development. This version was first tested internally by peers within our university to identify any obvious problems. Ten subjects agreed to test the video games and complete the questionnaire. Seven of the ten subjects were female, six had Type 1 insulin dependent diabetes mellitus (IDDM), four had Type 2 diabetes, and the average age of the group of testers was 38 years.

Two of the recruited subjects later declined to fully participate in testing the games, resulting in eight beta testers. Participation reminder emails were sent to these subjects, and it is not known why they declined to participate. Aside from participation reminder emails, no other subject contact was made following initial recruitment. Furthermore, remote analytics were unavailable for one tester, presumably because of a lack of device support or because the user turned off their data service. Although we limited the study to individuals who owned a supported device, it is possible that a model-carrier

Table 1 Data collected remotely from seven players of each game

	Hangman	QuizShow	CountDown	All
Total number of questions	120	173	109	402
Number of games started	100	51	31	182
Questions correct (%)	71 (95% CI 62–78)	71 (95% CI 64–77)	60 (95% CI 50–68)	68 (95% CI 63–72)
Games completed (%)	100 (95% CI 97–100)	53 (95% CI 46–60)	58 (95% CI 48–67)	80 (95% CI 76–84)
Total time played (min)	59	50	28	137
Time per question (s)	35	59	54	45

combination may not be supported. A summary of the remote testing and questionnaire results is given in Tables 1 and 2.

According to the metrics, the testers averaged 45 seconds per question and answered an average of 50 total nutrition questions each (402 questions/8 testers). Testers appeared to prefer Quizshow over playing the other games, i.e. testers attempted more QuizShow questions than other game questions. QuizShow also had the greatest length of time spent playing per question. Testers spent the greatest overall length of time playing Hangman and Quizshow. However, Quizshow also had the least number of games that were played to completion. Hangman had the most correct responses, perhaps indicating that it was the easiest.

Testers self-reported playing the game for 10–30 min (Table 3). This coincides with the measured metrics of the game. Seven of the eight testers reported somewhat enjoying the game play, and one tester reported enjoying the game play very much. Likewise, seven testers felt that the nutritional component of the games was somewhat challenging and one felt the nutritional component was very challenging. With one exception, all testers reported learning about nutrition while playing the games. All testers felt the nutritional information was relevant to them. Two testers felt it was very relevant and the remaining six felt the nutritional information was somewhat relevant.

Table 2 Players' perceptions of the games

Experience	Response	*n*
Game enjoyment	Yes, very much	1
	Yes, somewhat	7
	No	0
Challenging questions	Yes, very much	1
	Yes, somewhat	7
	No	0
Learned nutrition	Yes, very much	0
	Yes, somewhat	7
	No	1
Information relevance	Yes, very much	2
	Yes, somewhat	6
	No	0
Game difficulty	Too easy	4
	About right	2
	Too difficult	2
Game simplicity	Too simple	1
	About right	7
	Too complicated	0
Confusing game play	No	7
	Yes, some parts	1
	Yes, many parts	0

Table 3 Self-reported time spent testing games

Time played	No. of testers
1–2 hours	2
30 min-1 h	2
10–30 min	4

The questionnaire revealed additional insight into improvements in the design aspects of the game as well as game play and practical debugging of the user interface. Testers suggested improvements that included adding a greater variety of games and questions within games, allowing players to select their favourite food items for inclusion in the games, and additional hints to help the player find the correct answer. Bugs in the user interface and game play improvements in elements such as scoring were identified. In general, the testers remarked that the games were good and perhaps fun.

When asked 'What was your favourite aspect of the games?', testers commented on aspects of the games including enjoyment ($n = 2$), difficulty level ($n = 4$), variety of games and foods ($n = 1$), and the relevance of diabetes-related questions ($n = 4$).

The reported least favourite aspects of the game suggested that improvements could be made on the overall game play ($n = 3$), scoring mechanism ($n = 1$) and also the difficulty ($n = 3$).

Discussion

We created three simple video games to teach nutrition estimation and food comparison skills using formative research and user-centred design. The results of our formative research suggested improvements to the design of the mobile games, such as the appropriate level of difficulty, nutritional content, scoring systems and the acceptable game play duration on mobile phones. The results of our remote usability testing helped to further refine the user interface and game play.

Conventional laboratory-based usability testing can be expensive, time-consuming and unrealistic. In contrast, we found remote testing to be relatively inexpensive, rapid and capable of providing a real-world testing environment. Most of the online venues used for recruitment were free, although we spent about US$50 on two advertisements in local newspapers. Remote usability testing can effectively gather content-related issues, but device-related problems are more difficult.[19] For example, in our study, one of the participant devices did not transmit usage data to the server.

Some effort was required to identify appropriate recruitment venues, and then post messages regarding participation. We identified more than 15 message boards and other venues for potential subject recruitment, but many venues had policies against unsolicited postings. As a result we used eight websites, and followed up with updated postings several days per week for two weeks. We received a total of twelve responses, with two testers excluded because they did not have a supported mobile platform.

There was also an unexpected communication burden to recruiting and conducting tests remotely. Although the motivation behind remote testing included efficiently recruiting and conducting testing, some users required significant informational interaction and support via phone and email during testing. For example, one tester required 'talk through' assistance for installing and playing the games. Furthermore, 'in house' user testing may have fewer testers drop out of testing, particularly if done in one session. Our attrition rate over the one-week period was 20% (2/10).

Although some users required additional help installing the games on their devices, little effort was required by the researchers to facilitate the testing for the majority of the testers. Furthermore, self-reported survey data such as time spent playing games coincided with the collected metric data. This corroborates earlier research findings that self-reported user testing data can be as reliable as observed data.

There were several limitations to the present study. First, the samples sizes were very small. Second, although the remote metrics corroborated some of the self-reported data, the results could be biased. Finally, by combining subjects with type 1 and type 2 diabetes in a single study, important differences between the two groups may have been masked. However, as formative research, our study was illuminating. The knowledge gained about our target group was used to improve the effectiveness of our games. In addition, we learned several important lessons about the process of remote usability testing with mobile devices.

Remote usability testing methods pose new challenges involving recruitment venues and subject retention. These challenges notwithstanding, remote usability testing should be recognized as a valuable adjunct to laboratory testing due to advantages in cost, efficiency and context. As mobile data networks mature and devices continue to evolve, these methods are likely to gain even more prominence in the future.

Acknowledgements: Funding was provided by the Robert Wood Johnson Foundation/National Library of Medicine Public Health Informatics Training Program (3 T15 LM007442–06S1) and a grant from Robert Wood Johnson Foundation's Health Games Research program.

References

1 Norris SL, Lau J, Smith SJ, Schmid CH, Engelgau MM. Self-management education for adults with type 2 diabetes: a meta-analysis of the effect on glycemic control. *Diabetes Care* 2002;**25**:1159–71
2 Norris SL, Engelgau MM, Narayan KM. Effectiveness of self-management training in type 2 diabetes: a systematic review of randomized controlled trials. *Diabetes Care* 2001;**24**:561–87
3 Deakin T, McShane CE, Cade JE, Williams RD. Group based training for self-management strategies in people with type 2 diabetes mellitus. *Cochrane Database Syst Rev* 2005;CD003417
4 Wagner EH, Sandhu N, Newton KM, McCulloch DK, Ramsey SD, Grothaus LC. Effect of improved glycemic control on health care costs and utilization. *JAMA* 2001;**285**:182–9
5 Keers JC, Groen H, Sluiter WJ, Bouma J, Links TP. Cost and benefits of a multidisciplinary intensive diabetes education programme. *J Eval Clin Pract* 2005;**11**:293–303
6 Nagelkerk J, Reick K, Meengs L. Perceived barriers and effective strategies to diabetes self-management. *J Adv Nurs* 2006;**54**:151–8
7 Baranowski T, Buday R, Thompson DI, Baranowski J. Playing for real: video games and stories for health-related behavior change. *Am J Prev Med* 2008;**34**:74–82
8 AOL Games The Associated Press Poll. *Majority of Kids are Playing Video Games, But Not with Their Parents* 2007. See http://www.businesswire.com/news/google/20071112006305/en (last checked 23 January 2010)
9 Ryan R, Rigby C, Przybylski A. The motivational pull of video games: a self-determination theory approach. *Motiv Emot* 2006;**30**:344–60
10 Kreuter MW, Skinner CS. Tailoring: what's in a name? *Health Educ Res* 2000;**15**:1–4
11 Harris Interactive. *Teenagers: A Generation Unplugged*. See http://www.ctia.org/advocacy/research/index.cfm/AID/11483 (last checked 1 March 2010)
12 Harris L, DeShazo J, Pratt W. Diabetes and obesity: can video games help? In: Hayes BM, Aspray W, eds. *Health Informatics: A Patient-Centered Approach to Diabetes*. Cambridge, MA: MIT Press, 2010:238–71
13 USDA Food and Nutrient Database for Dietary Studies. *Agricultural Research Service, Food Surveys Research Group*. Beltsville, MD: USDA, 2004
14 Shalom MF. Making educational computer games "educational". Proceedings of the *2005 Conference on Interaction Design and Children*. Boulder, Colorado. 8–10 June 2005:56–61
15 Tzu-Hua W. Web-based quiz-game-like formative assessment: development and evaluation. *Computer and Education* 2008;**51**:1247–63
16 Lepper MR, Cordova DI. A desire to be taught: instructional consequences of intrinsic motivation. *Motiv Emot* 1992;**16**:187–208
17 Mensing CR, Norris SL. Group education in diabetes: effectiveness and implementation. *Diabetes Spectrum* 2003;**16**:96–103
18 Morten Sieker A, Henrik Villemann N, Simon Ormholt S, der Jan S. What happened to remote usability testing?: an empirical study of three methods. Proceedings of the *SIGCHI Conference on Human Factors in Computing Systems*. San Jose, California. 28 April–3 May 2007
19 Waterson S, Landay JA, Matthews T. In the lab and out in the wild: remote web usability testing for mobile devices. *CHI '02 Conference on Human Factors in Computing Systems*. Minneapolis MN. 20–25 April 2002