

THE DEVELOPMENT OF AN INSTRUMENT TO MEASURE MOBILE GAME QUALITY

Clarry Shchiglik
AAPT Telecom
680 George Street
Sydney
NSW 2000
Australia

Stuart J. Barnes*
Kent Business School
University of Kent
Medway Building
Chatham Maritime
Kent ME4 4AG
United Kingdom
Tel: +44-1634-888839

Eusebio Scornavacca
Department of Information Systems and Decision Science
University of Baltimore
1420 N. Charles St.
Baltimore
MD 21201
USA
Tel: +1-410-837-6624

klarik@xtra.co.nz, s.j.barnes@kent.ac.uk, escornavacca@ubalt.edu

Note: * Corresponding author

THE DEVELOPMENT OF AN INSTRUMENT TO MEASURE MOBILE GAME QUALITY

Clarry Shchiglik

Stuart J. Barnes

Eusebio Scornavacca

ABSTRACT

Electronic games are often considered one of the main consumer markets in the mobile commerce arena. This is an area that has thus far been neglected by literature. The research performed in this paper develops and validates the usefulness of an instrument that evaluates the quality of mobile games. Data collected by the instrument is grounded in subjective impressions, which can be applied in quantitative analysis for the production of mobile game metrics including the Mobile Game Quality Index. These objectives are carried out through a three-step methodology. First, an instrument to measure the construct of interest, mobile game quality is developed. Then the instrument's validity is examined and demonstrated in two phases of testing on mobile games. A five dimension structure of mobile game quality is revealed along with a validated 23-item instrument. The resulting quality model is found to have strong theoretical support in the literature.

Keywords: m-commerce; wireless; mobile; games; quality; evaluation.

INTRODUCTION

Games have been played since early civilization [51]. They have become a pervasive component of culture and are used to serve a wide variety of purposes including entertainment, education and the refinement of skills [12]. The influence of computers to games has been particularly dramatic. The introduction of computers to games approximately 40 years ago has broadened the way in which games can be played and inspired the emergence of a growing industry. In 2012, the global market size of the computer games industry was measured at US\$63 billion and it is forecasted to increase in size to US\$78 billion in 2017, a growth of approximately 24 percent [17].

The emergence of mobile devices has now introduced a new dimension to games and promises greater market scope [5]. The explosive growth of mobile phones has led to 6.8 billion subscribers worldwide, with 2 billion now using the faster 3G or 4G network standards [28]. Moreover, the mobility and network capability offered by mobile devices opens a raft of capabilities beyond boundaries of wired connectivity [50; 53]. Already the growth of games played over mobile phones (referred to as mobile games from here onwards) is a worldwide phenomenon and is predicted to grow from US\$5.4 billion in 2010 to US\$11.4 billion in 2014 [23]. With the proliferation of smartphones and advent of the Apple App Store and Google Play, the diffusion of mobile games has skyrocketed. According to Gartner, App downloads were expected to reach 102 billion in 2013 [24].

Experience in the gaming market has shown that while a game's brand may initially be able to attract consumers, it will not guarantee its success. Rather, the long-run success of a game is largely determined by its quality as perceived by consumers [26; 29]. In addition, games delivered over mobile devices operate in a different paradigm to those of other technologies, as dictated by

differences in infrastructure and user behavior. Interestingly, a good part of the problem with the initial wave of unsuccessful mobile games were due to the lack of understanding of quality from the perspective of the consumer and how their needs and expectations could be met over the mobile medium.

Consequently, just as with any other mobile service, it is fundamental to have an appreciation and ability to measure customer perceived quality in order to achieve successful deployment of these types of games [6; 29]. Advancement in the understanding of customer perceived mobile game quality, and the ability to evaluate such for any given mobile game, will provide an increased understanding of customer's adoption behavior and an ability to assess the likely success of commercialization.

Accordingly, there exist several studies dedicated to understanding m-commerce quality [4; 31]. These research efforts represent a sound beginning for the study of quality in the m-commerce domain. However, they do not strictly focus on the intricacies of mobile games. By focusing on mobile games, distinctive features that would otherwise be unobserved through broad m-commerce research can be exposed. Moreover, only two of these studies have gone beyond initial scoping and delimiting of the field and specifically sought to establish an instrument for the evaluation of customer perceived m-commerce quality [4].

The formal measurement of mobile game quality is similarly deficient within computer game literature, which has typically sought to focus largely on Playtesting [21; 48; 49] and a smaller amount is devoted to 'discount methods' such as expert review [39; 45]. Very few efforts have sought to formally assess game quality from surveys in a systematic, statistically valid and reliable way. If developed, such methods could provide a practical method of evaluating and comparing the quality of multiple games. This area of investigation provides another frontier of 'discount methods' for mobile game evaluation. Several research efforts can be identified as featuring game quality instruments [22]. However, in addition to their scarcity, a common element amongst these studies is a lack of disclosure of the actual instrument featured in the study. Moreover, while the instruments are grounded within literature, no extensive survey testing is performed to assert their validity for the measurement of game quality. Thus, it appears to date, academic literature has never disclosed a game-related quality instrument. Additionally, the identified game quality instruments are not specifically designed for mobile games. Yet, the various mediums over which games can be played possess distinct qualities [12].

Given the importance for research into the area of mobile game quality and, at the same time, the deficiency of literature available at present in this field, the objective of the current study is to develop a customer perceived instrument for the measurement of mobile game quality and then to demonstrate the validity and usefulness of this instrument. The final instrument is juxtaposed with the literature to add theoretical strength. This paper makes a contribution to the assessment of mobile game quality. In addition, the creation of an instrument to measure demand-side qualities could contribute to the mobile game development process, whereby quality is more formally scrutinized and matched with supply-side enhancements to engineer overall improvements in mobile game quality by targeting resources in specific areas of mobile game development. Thus, the instrument may be of value in quality function deployment (QFD) and similar quality engineering initiatives [53].

The remainder of the paper is structured as follows. The next section provides an overview of the research approach and design. This is followed by three sections that detail the results of the research within the three research steps. Note that, by design, details on the literature are integrated within appropriate locations in the paper's structure. The paper concludes with a discussion of the key research findings, limitations, and suggestions for further research and practice.

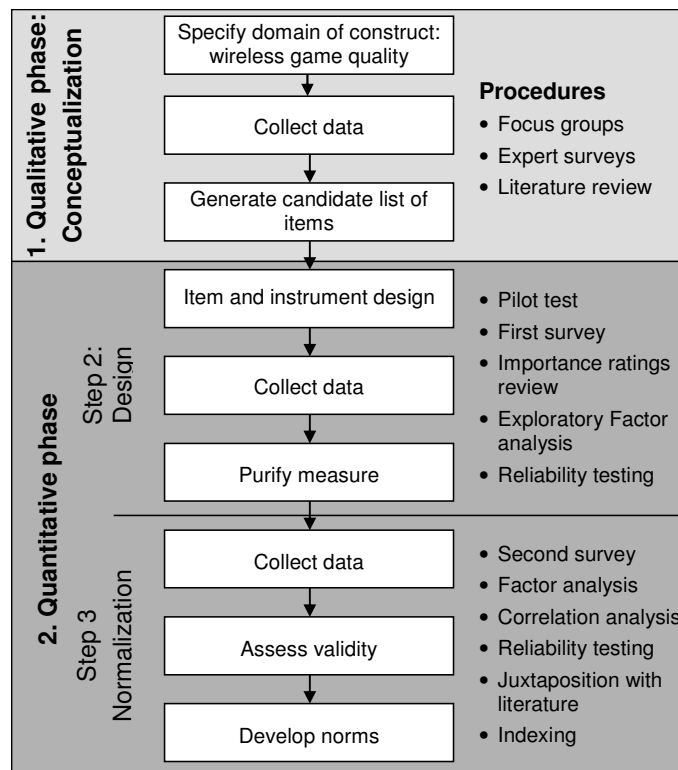
RESEARCH APPROACH

A number of researchers have studied and proposed procedural models for the development of instruments with better measures [2; 10; 55]. Over time, these suggested procedures have grown to become commonly applied in the Information Systems field [18; 1]. Aladwani and Palvia [1] have identified three generic common steps amongst many of these of procedural models: 1. Conceptualization; 2. Design; and 3. Normalization.

The research framework of the current study, shown in Figure 1, features two phases that summarize the three generic steps of instrument development identified by Aladwani and Palvia [1]. The first phase involved qualitative methods and comprised of the conceptualization step. This step aims to develop a candidate list of items based on a specified construct of interest and thus content validity is of key concern. Content validity was ensured by adopting a thorough approach involving three processes: a literature review, focus groups and an expert review. Conceptualization commenced by referring to literature to expressly specify the mobile game quality construct. Then with consideration of the defined construct, data collected from four focus groups was analyzed and combined with literature to produce an initial list of candidate items. This initial list of candidate items was then reviewed by experts.

The second phase involved quantitative methods and involved two steps: design followed by normalization. The purpose of the design step is to refine the list of candidate items put forward from the previous step and derive an initial instrument. This involves instrument design and performing a series of analytical procedures on data collected from pilot tests and a survey. Construct validity and reliability are considered throughout the design step. Normalization then sought to verify the instrument's validity and demonstrate its usefulness by establishing norms. These procedures were carried out by performing analytical procedures on data collected from a subsequent survey.

Figure 1: Research Framework



STEP I: CONCEPTUALIZATION

To begin with, the domain of the construct of mobile game quality was specified. This was carried out in consultation with literature. For the purposes of this research, mobile game quality is defined as a customer's evaluation of a mobile game satisfying their expectations. Previous attempts to define mobile game quality were not found. As a result, this definition is essentially a merger between related concepts within the fields of mobile games and quality [27; 43].

The next step of conceptualization was to generate items that capture the domain of the construct of interest: mobile game quality. In order to ensure content validity it is essential that the generated items are based on the specified domain [10]. Procedures that are beneficial for ensuring content validity are a review of literature and exploratory data collection, both of which were undertaken in this study [35].

A wide range of sources were reviewed in order to find literature relating to mobile game quality. This included journals, conference proceedings and postings on online community websites that concentrate on closely related areas. In addition, two types of exploratory data collection were performed; focus groups and an expert review.

Four focus groups were carried out, each consisting of six participants. In line with recommended procedure, participants were selected and grouped according to their previous experience with the Internet, computer games, mobile phones, and the mobile Internet in order to gain insight into consumer perceptions of mobile games from a variety of perspectives [38]. Although participants amongst all four groups were high Internet users, groups were unique with respect to participants' level of usage of the other three technologies: mobile Internet users (group A); computer game and mobile phone users (group B); mobile phone users (group C); and Internet only users (group D).

Each focus group ran for between 90 and 120 minutes and featured three stages: (1) Playing mobile games; (2) Focus group discussion; and (3) a Quality workshop.

Playing mobile games

Participants were provided with approximately 45 minutes to play and build familiarity with three designated games. To gain understanding of consumer perceptions with respect to a range of mobile games, each focus group was designated a unique set of three games consisting of one game from each of the following genres: multiplayer; role player; and classic. This interaction was recorded onto digital videocassette. A total of 12 games were played, represented by a unique set of three games for each of the four focus groups.

Focus group discussion

Once participants had played the three designated games they were asked open-ended questions requiring them to draw upon their experience. A conscious effort was made to ensure questions were clearly formulated, neutral, appropriately sequenced and easily understood. Focus groups were recorded on cassette tape. Indicative questions used at each focus group include:

- What problems did you have?
- What did you find easy?
- What did you like?
- What did you not like?
- Would you play the games again? If yes, why? If not, why?

Furthermore, follow-on questions evolved from several sources:

- Drilldown questions on specific issues expressed by participants to clarify thought and gain a deeper understanding.
- Comments provided by participants, which fuelled questions on related issues to allow a broader comprehension.
- Observations from participants playing mobile games in stage one of the focus group.

Questions were not preconceived for specific focus groups. However, a number of distinct follow-on questions arose amongst focus groups. Most noticeably, groups consisting of experienced game players (groups A and B) raised considerable concerns regarding a lack of game functionality and media richness, while groups consisting of inexperienced game players (groups C and D) tended to raise issues concerning perceptions of control and challenge in game-play.

Quality workshop

For the final stage of the focus group, a quality workshop was run to uncover perceived qualities of an excellent mobile game. Bossert [8] recommends a three-stage process for such workshops. Participants were first provided with post-it notes and asked to answer the following question: “What are the qualities of an excellent mobile game?” Participants were instructed to work in silence, write one quality per post-it note and encouraged to also write a brief explanation and rationale for the proposed quality.

Once all participants felt they had exhausted their ideas, participants were then sorted into two groups of three to combine their post-it notes into affinity groups. Participants initially performed this task in silence, moving post-it notes around and creating headings as felt appropriate. Finally, participants worked as one group to develop one combined affinity group of demanded qualities.

Data collected from focus groups were analyzed using grounded theory techniques [56]. Initially, data gathered from stage one, user behavior, and stage two, focus group discussion, were organized. First and foremost, the observations concerning the qualities of an excellent mobile game were coded openly and qualitatively. Following this, these observations were re-coded axially, converging into categories: user friendliness, media richness, interactivity, price, rewards, responsiveness, functionality, multiplayer, personalization and enjoyment.

Literature was compared and contrasted with findings from the four focus groups to produce a total of 40 items conceptualizing mobile game quality. The initial list of 40 candidate items was then sent out to 12 experts for feedback to ensure the mobile game quality construct was adequately captured. Experts were either in the category of academic or industry, and were involved in either game research or development. Their knowledge offered an additional layer of insight into mobile game quality and, as a result, further enhanced content validity. From the 12 experts contacted, only eight responses were received. Three were from academia and five from industry. Once all responses had been collected, expert’s comments were reviewed by researchers involved in the current study. This process resulted in seven changes to the instrument. Changes comprised of the addition of three items, the amendment of one item and the removal of three items.

STEP II: DESIGN

In practice, it is not feasible to administer all possible items that could potentially be used to measure mobile game quality nor the entire list of candidate items put forward from the conceptualization step [10]. For this reason, it is necessary to purify the candidate list of items put forward from the conceptualization step in order to retain a sample of items that represent the construct of mobile game quality and offer a highly correlated measure of mobile game quality. In addition, it is possible an initial instrument or items generated from conceptualization may present

respondents with operational issues. These operational issues are identified and resolved and purification is achieved in the current study with the application of analytical procedures of data collected from pilot testing and a survey.

Before pilot tests commenced, which customer perspectives would be measured needed to be addressed. This question arises as service quality has been previously operationalized with several different measurement approaches. Literature has sought to assess service quality with measurement of customer perceptions (i.e., service quality as perceived by a customer), customer expectations (i.e., expectations of the service quality that will be delivered), and importance ratings (i.e., the importance of a quality to a customer).

The approach applied to the current study is to measure customer perceptions and importance ratings. Customer perceptions consider the degree to which a quality has been attained, while importance ratings consider the importance of a particular quality. This approach is believed to capture the most important aspects of quality and at the same time keep the number of assessments that need to be made to a manageable number [4]. Furthermore, Mazis et al. [37] suggest the combined measurement of customer perceptions and importance ratings is the most efficient approach to use where the objective is to predict behavioral intention or actual behavior.

A pilot test was carried out to identify issues with the instrument and overall survey procedure. Eleven undergraduate students participated in the pilot test. Each participant was asked to play a mobile game for a period of time that they felt was sufficient to be able to then assess its quality. Feedback was received from participants through a questionnaire, which included the developed instrument, and also a one-to-one, face-to-face interview.

Pilot tests revealed detailed insights into the instrument and survey process. Participants felt there were an excessive number of items and several items had either interpretation difficulty or questionable applicability to the assessment of mobile game quality. Feedback received from the pilot tests resulted in the rephrasing of three items and removal of five items. Thus, the instrument was reduced to 35 items.

A survey then was performed in order to allow the instrument to be refined. The survey was carried out by evaluating three mobile games, as they were the most common mobile game format available at the time and students were targeted as respondents as they are known to be a key market for mobile games. To increase the response rate, respondents were asked to evaluate only one game and were offered the opportunity to win one of two NZ\$50 (US\$42) vouchers. This resulted in 86 responses, 85 of which were usable.

The demographics of the sample were as follows. Two-thirds of respondents were male and one-third female. The majority of the sample lay in the age brackets 16-20 (32%) and 21-25 (52%). Also, an overwhelming majority of the sample (82%) had previously played a mobile game.

Data collected relating to importance ratings offered the first means by which the instrument could be refined. As importance ratings measure the relevance of an item to the assessment of mobile game quality, they allow the instrument to be restricted to contain only salient items. Thus, one approach employed to ensure the final instrument contained items considered most relevant for the measurement of mobile game quality was the removal of items with a low importance rating. This was carried out in a systematic process whereby items that fell into a category with the significantly lowest importance rating means were removed. This resulted in the removal of four items.

Exploratory factor analysis was then applied to the collected data to reveal the underlying factors of mobile game quality. Items removed due to low importance ratings were included in this analysis to help identify these underlying factors. Factor analysis was carried out based on two key criteria. Firstly, to retain items that loaded greater than 0.55 on a single factor and less than 0.5 on all other factors [25]. Secondly, to retain factors that were found to consist of at least two items.

Factor loadings represent correlations between original item scores and factors. Thus, convergent validity is claimed when instrument items load highly on relevant factors. Conversely, discriminant validity is shown when items load more highly on one factor than on others. Factor analysis also reveals sub-constructs of the overall construct of interest.

The Varimax factor rotation converged in 15 iterations and revealed eight factors. Three of these eight factors contained only one item. Such factors are common for factor analysis procedures performed at an early stage of an analytical process and the items within these factors are not considered to have a significant relationship with the focal construct of interest, i.e., mobile game quality [10]. For this reason, lone item factors were eliminated along with the items they represented. Furthermore, seven items were found to not have a factor loading over the 0.55 cut-off for any factor and were removed from the instrument. However, one of these items had already been removed due to low importance rating; thus, factor analysis resulted in reducing the instrument by nine items.

Factor analysis revealed five dimensions of mobile game quality that are distinct and uni-dimensional and thus, indicate evidence of both convergent and discriminant validity. These factors were: Ease of use, Content quality, Responsiveness, Aesthetic appeal and Gaming experience. Explanations of the dimensions of mobile game quality are as follows:

- *Ease of use*: the ease of learning, playing and understanding the objective of a game. The dimension also considers ease with which it is possible to navigate.
- *Ease of access to quality content*: the straightforwardness with which accurate, relevant and easy to understand content can be accessed.
- *Responsiveness*: the period of time taken to receive a response and the timeliness of information received.
- *Aesthetic appeal*: the attraction of a game base on the image it conveys and its use of multimedia.
- *Gaming experience*: qualities concerning flow and enjoyment from playing games, e.g., keeping a user's attention focused, the incentive of playing and the interest gained from the experience.

Reliability was the final analytical procedure of the design step. Cronbach's α was 0.934 for the overall instrument and ranged from 0.719 to 0.919 for measurable dimensions. All α coefficients are in the range of acceptability [42].

As a whole, procedures undertaken in the design step demonstrated convergent and discriminant validity as well as reliability. The step also resulted in the removal of 18 items (pilot tests five items, importance ratings four items and factor analysis nine items), rephrasing of three items and subsequent addition of four items. Thus, the instrument was refined to contain 26 items.

STEP III: NORMALIZATION

Normalization was the third and final step of the research framework. A key focus of the step is construct validity. Normalization involves a subsequent independent verification and validation of the instrument's construct validity as initially proposed previously in the design step. In addition, normalization sets out to demonstrate the usefulness of the instrument by establishing norms and by juxtaposition of the instrument with the salient theoretical and conceptual literature. These procedures are achieved by analyzing data from a second survey.

Second application of the survey

The second survey was carried out by evaluating three anonymous mobile games: Alpha, Beta and Gamma. Alpha is a multiplayer game designed to test memory, the purpose of Beta is to solve word puzzles, and Gamma is a multiplayer, multiple-choice general knowledge game where correct answers push the player along a racetrack. The survey used a five-point Likert scale where

the anchors range from 1 (“strongly disagree”) up to 5 (“strongly agree”). In addition, to encourage participation and the quality of response, a prize draw with a NZ\$250 (US\$210 approx.) first prize and NZ\$50 (US\$42 approx.) second prize was offered. These procedures resulted in 127 responses, of which 125 were usable.

Approximately two-thirds of respondents were male. The sample consisted primarily of respondents in the 21 to 25 (36%) and 26 to 30 (52%) age groups. In addition, the majority of the sample (83%) had experience with playing a mobile game. Importance ratings and perception scores were expressed in terms of their mean and standard deviation values.

The Varimax factor rotation converged in nine iterations and confirmed a five factor structure of mobile game quality, namely: ease of use, content quality, responsiveness, aesthetic appeal and gaming experience. The results are shown in Table 1.

Table 1: Exploratory factor analysis (loadings of 0.5 and above, second survey)

No.	Item Description	Factor				
		<i>Ease of use</i>	<i>Content quality</i>	<i>Responsiveness</i>	<i>Gaming experience</i>	<i>Aesthetic appeal</i>
1	Easy to use	0.792				
2	Easy to learn and operate	0.820				
3	Knew what I had to do	0.808				
4	Easy to navigate	0.678				
5	It is easy to find things in the game	0.628				
6	Provides accurate content		0.732			
7	Provides relevant content		0.797			
8	Provides easy to understand content		0.624			
9	<i>Provides consistent navigation</i>			0.529		
10	<i>Provides fast navigation to what I intend to find</i>			0.569		
11	Provides timely content			0.630		
12	Has a fast response time			0.849		
13	Provides feedback quickly			0.824		
14	Meets my expectations				0.573	
15	Provides an enjoyable experience				0.694	
16	Provides value for money				0.562	
17	Keeps my attention focused				0.631	
18	<i>Provides an appropriate level of challenge</i>				0.522	
19	Provides an incentive to play				0.647	
20	Benefits are gained from playing with others				0.604	
21	Has innovative game play				0.725	
22	Interesting to play				0.629	
23	Has visually pleasing graphics					0.763
24	Looks professionally designed					0.834
25	Has an attractive appearance					0.849
26	Design is suited for the mobile phone					0.616

Items 9 and 18, both failed to produce a factor loading greater than 0.55 for any factor. Item 10, on the other hand, loaded above the 0.55 cut-off for a different factor than previously identified. In the earlier factor analysis, item 10 loaded 0.552 for the factor content quality, whereas in this factor analysis the loading was 0.569 for responsiveness. There is the possibility that the exploratory nature of the first factor analysis caused the item to be initially and erroneously placed into the content quality factor. Further research is required to confirm this supposition. However, in any case, the suggestion of a lack of discriminant validity of this item was sufficient to warrant its removal from the instrument. Moreover, in the earlier factor analysis this item had already indicated a discriminant validity concern, but was retained as part of the content quality factor as the discrepancy was only marginal. A closer inspection of the item's phrasing, "fast navigation to what I intend to find", further supports these empirical findings; "fast" is indicative of responsiveness, while "navigation to what I intend to find" would likely relate to easy access to content.

The inability of item 18 to meet the 0.55 loading cut-off is particularly interesting as challenge is considered within the flow literature to be an integral part of the interactive experience [13]. However, perhaps the term 'challenge' offers respondents an unintended interpretation that inspires connotations of work, effort and exercise – all of which are incompatible with the factor "Gaming experience," the factor which had earlier encompassed the item. With this premise, the removal of item 18 reduces the potential ambiguity of the instrument. Further explanation of the absence of item 18 from the gaming experience factor can be gained from considering the scope of item 22, "Interesting to play." The two items are in many ways related. Moreover, a review of the scope of item 22 reveals it encapsulates the intended meaning of item 18.

Flow literature asserts that if a game is not challenging it will become boring quickly; conversely, if it is too challenging a player will feel frustrated and discouraged [36]. Comments received from the exploratory data collection procedures of the conceptualization step as well as the first and second surveys included: "Boring game, not challenging, or interesting and no graphics" (first survey), "Pretty straight forward game that does not capture people's interest. No challenge" (first survey) and "...the [trivia] questions were too hard." (focus group discussion). The comments present a strong association between the concepts of challenge and whether a game is interesting. They also express respondents' dissatisfaction with games that are either overly easy or difficult. As a result, the inclusion of item 22, an item that determines whether a game is interesting to play, acts to effectively include the intended qualities of 'challenge' into the gaming experience factor.

The 23 items retained in the final instrument reveal five factors of mobile game quality that are distinct, uni-dimensional and verify findings from the design step. Cronbach's α was calculated for each of the five underlying dimensions of mobile game quality. Table 2 shows the calculated Cronbach's α coefficients: for the overall instrument the value was 0.932 and for the five dimensions it ranged from 0.823 to 0.888. All of these α coefficients are in the range of acceptability [42]. Moreover, Straub and Carlsson [55] recommend a strict cut-off α coefficient of 0.8 for the professional application of survey constructs, a criterion which is met by all dimensions in our study.

Table 2: Cronbach's α coefficients (second survey)

Dimension	Number of items	Cronbach's α
Ease of use	5	0.886
Content quality	3	0.833
Responsiveness	3	0.830
Gaming experience	8	0.888
Aesthetic appeal	4	0.830
Overall	23	0.932

Convergent validity is concerned with the extent to which multiple measures of the same construct agree with each other [9]. The correlation coefficient between the instrument and overall quality was found to be 0.771 with $p < .001$, demonstrating convergent validity. Discriminant validity refers to the extent to which measures of different constructs are distinct [9]. As we can see from Table 1, the factors are clear and distinct in their loadings, with no overlap.

To further confirm the validity of the final 23-item instrument, the dimensions were examined via a second-order model of mobile game quality using confirmatory factor analysis and partial least squares path modeling in Smart PLS [47]. All items loaded on their respective factors at $p < .001$ (t-values ranging from 6.447 to 53.614). Further, all factors were strongly related to mobile game quality at $p < .001$ (path coefficients ranging from 0.181 to 0.398; t values ranging from 9.563 to 15.363).

Further validity and reliability data are reported in Table 3. Convergent validity was confirmed by average variance extracted (AVE) values, which ranged from 0.564 to 0.747, above the 0.5 level recommended by Fornell and Larcker [19]. Similarly, composite reliability ranged from 0.888 to 0.917, well above the recommended levels of 0.7 and 0.8 [42; 55]. Discriminant validity was tested by comparing the intercorrelations with the square-root of AVE. In all cases the square-root of AVE is well above intercorrelations in the associated rows and columns.

Juxtaposition of survey instrument with the literature

The final instrument for measuring mobile game quality is rooted in the 'voice of the customer or user'. Notwithstanding, from a juxtaposition of the instrument with the salient literature, it is also clear that parts of the instrument are grounded in core elements of information systems theory. Let us briefly consider these elements in order to confirm conceptual and theoretical support for the final instrument from the salient literature on information systems and e-commerce quality:

- *Ease of use.* The ease of use construct, often referred to as usability, has its roots in human-computer interaction theory [14; 15; 39]. This construct has also proven important in Web usability research [40; 41; 54] and more recently has been shown to be integral to assessments of Web quality [6; 7; 20; 59; 60]. Overall, there is very strong literature support for the ease of use items in the mobile game quality instrument.
- *Content quality.* The items in this construct refer to understandable, relevant and accurate information. These are core elements of information quality, which is a longstanding and important concept in information systems quality [3; 16; 58]. Information quality has also proven central to many instruments for Web quality evaluation [1; 6; 7; 33; 57].

Table 3: Intercorrelations between mobile game quality dimensions

Dimensions	AVE	CR	Aesthetic appeal	Ease of use	Game experience	Content quality	Responsiveness
<i>Aesthetic appeal</i>	0.669	0.888	0.818				
<i>Ease of use</i>	0.689	0.917	0.328	0.830			
<i>Game experience</i>	0.564	0.910	0.606	0.616	0.751		
<i>Content quality</i>	0.747	0.898	0.370	0.625	0.589	0.864	
<i>Responsiveness</i>	0.738	0.839	0.343	0.414	0.465	0.466	0.859

Note: AVE= average variance extracted; CR= composite reliability; $\sqrt{\text{AVE}}$ on diagonal. *Responsiveness*. These items refer to system responsiveness, in particular the speed of receiving content and feedback. Again system response time is typically considered a central element of information systems success [3; 16]. It is also found in many instruments for measuring Web quality [30] .

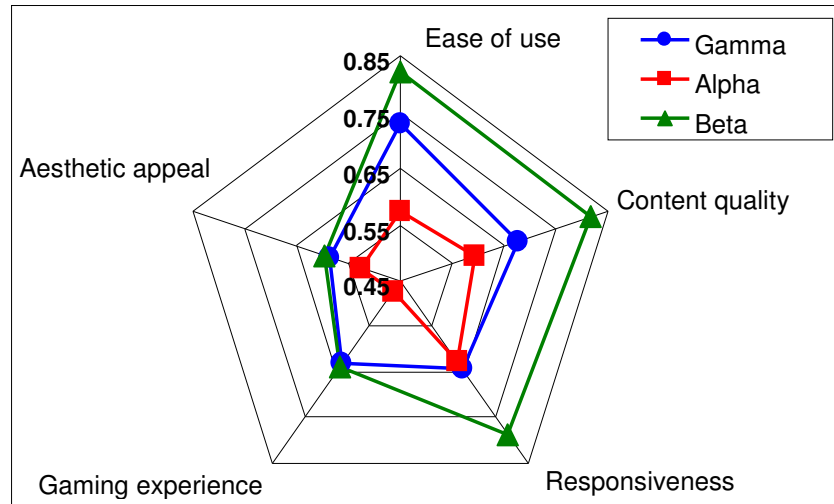
- *Gaming experience*. This construct is new and does not have a foundation in the information systems literature. The items in the construct are varied, but all of them refer to elements of valuation of the experience of game playing. As such, they tend to find support in the literature on heuristic evaluation of computer games [45] and mobile games. Moreover, the items are supported by the recent study of Penttinen et al. [44], who also have a construct for mobile game experience within the instrument that they develop.
- *Aesthetic appeal*. The final construct within the mobile game quality instrument refers to attractive, professional and visually pleasing graphics that are suited to the mobile device. This aspect of design has been shown to be a key dimension in a very large number of instruments for measuring Web quality [1; 11; 30; 32; 34; 59].

Demonstration of application of the instrument

The usefulness of the instrument was demonstrated by developing norms. A raw score obtained from an instrument measurement on its own is not particularly informative; however, establishing norms created a benchmark and thus allowed meaningful comparisons and interpretations of the quality of the three evaluated mobile games. Norms were developed by calculating an index, referred to in the current study as the Mobile Game Quality Index (MGQI). The MGQI is calculated by dividing the weighted score by the maximum score. The weighted score represents each respondent's perception score multiplied by the importance rating attached to it and the maximum score represents the mean importance rating multiplied by 5 – the maximum rating attainable. Overall, Beta is benchmarked highest with an overall MGQI of 0.72. Beta is relatively closely followed by Gamma, which has a MGQI of 0.66 and Alpha is well below the other two games with a MGQI of 0.55.

Further, more detailed, comparisons between the three mobile games can be made by applying the MGQI with respect to the five dimensions of mobile game quality, namely, ease of use, content quality, responsiveness, aesthetic appeal and gaming experience. The required calculations commenced by grouping data according to dimensions. Then the total score for each dimension was divided by its maximum score to derive a MGQI. The results of these calculations are shown as a graphical representation in Figure 2. These comparisons are particularly beneficial as they permit a closer examination into why some games fared better than others in terms of the overall MGQI. Note the scale have been restricted to 0.45 to 0.85 to help draw clearer comparisons.

Figure 2: Radar chart of MGQI dimensions with respect to the three evaluated games



Overall, Beta was found to have a substantially higher MGQI score than the other two games with respect to the dimensions of ease of use, content quality and responsiveness. However, in terms of the other two dimensions Beta scored similarly to Gamma. In addition, it was found that while Gamma scored substantially higher than Alpha at an overall level, the two games had a similar MGQI in regards to the responsiveness dimension.

The relatively high MGQI score achieved by Beta for the ease of use dimension reflects the game's uncomplicated nature. Beta simply requires players to read a riddle and then click to see the answer. There is no score keeping or game mechanism to determine whether the player was correct. Conversely, Gamma, for example, presents a more complicated task proposition. For example, a game cycle for Gamma requires typically upwards of a dozen screens to be downloaded and viewed compared to only two for Beta. Thus, by its very nature, Gamma is expected to be more difficult to use.

A plausible explanation for the similarity in MGQI scores between Gamma and Alpha for the responsiveness dimension rests with their game design. Both games are multiplayer and are thus heavily reliant upon network connectivity. Moreover, the way in which multiplayer gaming is implemented in both games is identical: players are pitted against one another to complete a race course and progress by correctly answering questions. They answer questions synchronously and independently of each other and thus their ability to compete against one another is heavily reliant upon the quality of their network connection, i.e., should a player receive favorable connectivity then they will be in an advantageous position.

Compared to Gamma and Alpha, Beta scores substantially higher for the responsiveness dimension. In terms of responsiveness qualities, Beta is most notably distinctive for being single-player and also excluding graphics other than what is seen on its main introduction screen (i.e., less data needs to be downloaded). As a result, Beta is comparably less reliant upon network connectivity. Results regarding responsiveness find multiplayer games disadvantaged due to their reliance upon mobile network connectivity. This reaffirms the need for game design to cater for limitations of mobile networks.

Another interesting observation regarding MGQI results concerned the aesthetic appeal dimension. This was the only dimension in which all three games score below 0.6 on the MGQI scale. It is believed that MGQI scores for this dimension are a reflection of the restricted capabilities of the format of the mobile games evaluated to cater for interactive multimedia capabilities that are typically suited for games such as sound and moving objects. Moreover, aesthetic appeal was noticeable for having the lowest spread of MGQI scores. For this dimension, MGQI scores ranged

from 0.53 to 0.60. It is possible that the three evaluated games bear many similarities in terms of aesthetic appeal. Alternately, it is also possible that the limited multimedia capability of some mobile devices presents little opportunity for games to differentiate themselves in terms of this dimension.

CONCLUSIONS AND FUTURE WORK

This paper aimed to develop and demonstrate the usefulness of a valid customer perceived instrument that specifically caters for the measurement of mobile game quality. Research was carried out in the form of a three-step, two-phase investigation. The first phase focused on the development of an instrument that conceptualized the construct of mobile game quality. The second phase focused on validity testing and demonstrated the usefulness of the developed instrument through a process design and then normalization.

Results of the two-phased investigation uncovered five dimensions of mobile game quality: ease of use; content quality; responsiveness; aesthetic appeal; and, gaming experience. The results provided evidence for the psychometric properties of the 23-item instrument to measure mobile game quality. A rigorous process to ensure content validity and tests for convergent and discriminant validity as well as reliability demonstrated evidence of a robust instrument.

Results from the study indicate the developed instrument is a useful diagnostic tool for the assessment of mobile game quality from a customer's perspective. The instrument derives a MGQI score, which can be applied at an overall mobile game quality level by using the entire 23-item instrument or at a specific mobile game quality dimension level, i.e., using a sub-scale of one of the five dimensions of mobile game quality. Furthermore, the instrument can be applied either as a cross-sectional survey to provide a benchmark against other mobile games or longitudinally to ascertain the consequences of amendments and the affect of time. For industry, the developed instrument serves as a guide to realize customer needs and wants and accordingly produce mobile games that effectively meet these requirements.

The instrument developed in this paper provides some of the raw material to initiate quality function deployment within the development process. Quality function deployment is a "structured and disciplined process that provides a means to identify and carry the voice of the customer through each stage of product and or service development and implementation" [53]. By using an instrument for measuring customer needs a mobile game developer will be able to identify and quantitatively assess priorities and match these to design requirements in a systematic way. The instrument provides the first step in identifying demand-side customer needs. By assessing development priorities (using tools as the analytical hierarchy process) and matching these to elements of subsystems and parts of the mobile game a supply-side engineering process can be enacted. For example, the heuristic "does the interface include scorekeeping" would intuitively be expected to have a strong link back to the quality "provides relevant content". Matrices of qualities and heuristics need to be developed in order for key relationships to be uncovered and, as a result, the voice of the customer to be deployed throughout mobile game design and development. It is not necessary, nor likely to be feasible, to model all plausible relationships. Rather, the aim is to understand the strong relationships in order to provide mobile game developers with critical guidance. The translation of customer demanded qualities into specific heuristics in such a fashion offers great value as it is ultimately the customer that determines the quality and success of a mobile game.

Key limitations of the research performed concerned focus groups and surveys. Both procedures exposed respondents to a restricted perspective of mobile games. In addition, both the first and second surveys' respondents experienced varying degrees of the evaluated mobile games before making their assessments. While some respondents had a winning experience others had a losing experience. Respondent's perceptions of a game are likely to have been affected by their

success. Moreover, in cases where respondents had lost, they may have only experienced a partial representation of a game, for example, not seeing the benefits that a game presents upon winning.

There are number of possibilities for future research to advance the progress made in the current study. This is particularly so with respect to the refinement and improvement of the developed instrument as “no good canvas is completed in a first attempt” [46, pp. 218]. The current study presents a first step in the development of a valid instrument that specifically caters for the measurement of mobile game quality and there remains scope for further improvement. A possibility for future research is to conduct further exploratory data collection for the purpose of understanding consumer perceptions of mobile game quality with respect to new mobile game standards which are constantly under development, calling for the need for further studies based on these new platforms.

ACKNOWLEDGMENTS

The support of New Zealand Telecom is gratefully acknowledged in this research.

REFERENCES

- [1] Aladwani, A.M. and Palvia, P.C. (2002). Developing and validating an instrument for measuring user-perceived web quality. *Information and Management* (39:6), 2002, pp. 467-476.
- [2] Bagozzi, R.P. *Causal Models in Marketing*, Wiley, New York, NY, 1980.
- [3] Bailey, J. E. and Pearson, S. W. Development of a tool measuring and analyzing computer user satisfaction. *Management Science* (29:5), 1983, pp. 530-544.
- [4] Barnes, S. J., Liu, K. and Vidgen, R. T. Evaluating WAP news sites: The WebQual/M approach. In Proceedings of the 9th European Conference on Information Systems. Bled, Slovenia, 2001.
- [5] Barnes S. J. Provision of services via the Wireless Application Protocol: A strategic perspective. *Electronic Markets* (12:1), 2002, pp. 14-21.
- [6] Barnes S. J. and Vidgen, R. T. An integrative approach to the assessment of e-commerce quality. *Journal of Electronic Commerce Research* (3:3), 2002, pp. 114-127.
- [7] Barnes S. J. and Vidgen R. T. Data triangulation and Web quality metrics: a case study in e-government. *Information and Management*, (43:6), 2006, pp. 767-777.
- [8] Bossert, J.L. *Quality Function Deployment, a practitioner's approach*, ASQC Quality Press, Wisconsin, 1991.
- [9] Campbell, D.T. and Fiske, D.W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin* (56), 1959, pp. 81-105.
- [10] Churchill, G. A. A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research* (16:1), 1979, pp. 64-73.
- [11] Cox, J. and Dale, B. G. Service quality and e-commerce: an exploratory analysis. *Managing Service Quality* (11:2), 2001, pp. 121-131.
- [12] Crawford, C. *The art of computer game design*. Washington State University, Vancouver. Available from: <http://www.vancouver.wsu.edu/fac/peabody/game-book/Coverpage.html> (accessed 7 October 2006), 1982.
- [13] Csikszentmihalyi, M. *Beyond Boredom and Anxiety*. Jossey-Bass, San Francisco, CA, 1975.
- [14] Davis, F. Perceived usefulness, perceived ease of use, and user acceptance of information technology, *MIS Quarterly* (13:3), 1989, pp. 340-391.
- [15] Davis, F. User Acceptance of Information Technology: System Characteristics, User Perceptions,

- and Behavioral Impacts, *International Journal of Man-Machine Studies* (38), 1993, pp. 475-487.
- [16] DeLone, W.H., and McLean, E.R. Information Systems Success: The Quest for the Dependent Variable, *Information Systems Research* (3:1), 1992, pp 60-95.
- [17] DFC Intelligence. *DFC Intelligence forecasts worldwide game market to reach \$79 billion by 2017*. Available from: <http://www.dfciint.com/wp/?p=353> (accessed September 24 2013), 2010.
- [18] Doll, W. J. and Torkzadeh, G. The measurement of end user computing satisfaction. *MIS Quarterly* (12:2), 1988, pp. 259-274.
- [19] Fornell, C. and Larcker, D. Evaluating structural equation models with unobservable variables and measurement error, *Journal of Marketing Research* (18:1), 1981, pp. 39–50.
- [20] Flavian, C., Guinalu, M. and Gurrea, R. The role played by perceived usability, satisfaction and consumer trust on website loyalty. *Information and Management* (43), 2006, pp. 1-14.
- [21] Fullerton, T., Swain, C., Hoffman, S. *Game Design Workshop: Designing, Prototyping, and Playtesting Games*. CMP Books, 2004.
- [22] Gao, Y. Appeal of online computer games: A user perspective. *The Electronic Library* (22:1), 2004, pp. 74-78.
- [23] Gartner. *Gartner says worldwide mobile gaming revenue to grow 19 percent in 2010*. Available from: <http://www.gartner.com/it/page.jsp?id=1370213> (accessed February 23, 2011), 2010.
- [24] Gartner. *Gartner Says Mobile App Stores Will See Annual Downloads Reach 102 Billion in 2013*. Available from : <http://www.gartner.com/newsroom/id/2592315> (accessed September 30 2013), 2013.
- [25] Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. *Multivariate data analysis with readings*. Prentice Hall, Englewood Cliffs, NJ, 1998.
- [26] Heydon, P. *Disruption = value creation in games. Presentation to the Edinburgh Interactive Festival, August 25, 2010*. Available from: <http://avistapartners.com/news-events.html> (accessed February 23, 2011), 2010.
- [27] Hoyer, R.W. and Hoyer, B.Y. (2001). What is quality? *Quality Progress* (7), 2001, pp. 53-62.
- [28] International Telecommunication Union. *The world in 2013: ICT facts and figures*. Available from: <http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2013.pdf> (accessed October 1, 2013), 2013.
- [29] Kangas, S. *Benchmarking Literature Review*, European Commission, Brussels, 2003.
- [30] Kim, S. and Stoel, L. Dimensional hierarchy of retail website quality. *Information and Management* (41:5), 2004, pp. 619-633.
- [31] Landor, P. *Understanding the foundation of mobile content quality: A presentation of a new research field*. In the proceedings of the 36th Hawaii International conference of System Sciences, 1-10. Waikoloa, HI, 2003.
- [32] Loiacono, E. T., Watson, R. T. and Goodhue, D. L. WebQual: A Web Site Quality Instrument. Working Paper 2000-126-0, University of Georgia, 2000.
- [33] Liu, C. and Arnett, K. Exploring the factors associated with Web site success in the context of electronic commerce. *Information and Management* (38:1), 2000, pp. 23-33.
- [34] Madu, C. N. and Madu, A. A. Dimensions of e-quality. *International Journal of Quality and Reliability Management* (19:3), 2002, pp. 246-58.
- [35] Malhotra, M. K. and Grover, V. An assessment of survey research in POM: From construct to theory. *Journal of Operations Management* (16:4), 1998, pp. 403–423.

- [36] Mallon, B. and Webb, B. Structure, causality, visibility and interaction: propositions for evaluating engagement in narrative multimedia. *International Journal of Human-Computer Studies* (53:2), 2000, pp. 269-287.
- [37] Mazis, M. B., Olli, T. A. and Kleppel, R. E. A comparison of four multi-attribute models in the prediction of consumer attitudes. *Journal of Consumer Research* (2), 1975, pp. 38-52.
- [38] Morgan, D. and Stinson, L. What are Focus Groups? American Statistical Association, Alexandria VA, 1997.
- [39] Nielsen, J. *Usability Engineering*, Morgan Kaufmann, CA, 1993.
- [40] Nielsen, J. User Interface Directions for the Web. *Communications of the ACM* (42:1), 1999, pp. 65-72.
- [41] Nielsen, J. *Designing Web Usability*, New Riders, IN, 2000.
- [42] Nunnally, J. C. *Psychometric theory*. McGraw-Hill, New York, NY, 1967.
- [43] Oliver, R. L. A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research* (17), 1980, pp. 460-469.
- [44] Penttinen, E., Rossi, M. and Tuunainen, V. (2010). Mobile games: analyzing the needs and values of consumers. *Journal of Information Technology Theory and Application* (11:1), 2010, pp. 5-22.
- [45] Pinelle, S., Wong, N., Stach, T. Heuristic evaluation for games: usability principles for video game design. In proc CHI, *ACM Press*, 2008, pp. 1453-1462
- [46] Pitt, L. F., Watson, R. T. and Kavan, C. B. Service quality: A measure of information systems effectiveness. *MIS Quarterly* (19:2), 1995, pp. 173-187.
- [47] Ringle, C. M., Wende, S., Will, A. SmartPLS 2.0 (beta), Available from <http://www.smartpls.de> (accessed 7 March 2011), 2005.
- [48] Rouse, R. *Game Design: Theory and Practice*. Wordware Publishing, 2001.
- [49] Schell, J. *The Art of Game Design*. Morgan Kaufmann, 2008.
- [50] Scornavacca, E., Barnes, S. J. and Huff, S. Mobile business research published in 2000-2004: Emergence, current status, and future opportunities. *Communications of the Association for Information Systems* (17), 2006, pp. 635-646.
- [51] Shaw, I. *The Oxford History of Ancient Egypt*. Oxford University Press, USA, 2004.
- [52] Siau, K. I. and Shen, Z. Mobile communications and mobile services. *International Journal of Mobile Communications* (1:2), 2003, pp. 3-14.
- [53] Slabey, R. QFD: a basic primer. Excerpts from the implementation manual for the three day QFD workshop. *Transactions from the Second Symposium on Quality Function Deployment*, Novi, MI, 1990.
- [54] Spool, J., Scanlon, T., Schroeder, W., Snyder, C. and DeAngelo, T. *Web Site Usability: A Designer's Guide*. Morgan Kaufmann, 1999.
- [55] Straub D. W. and Carlsson, C. L. Validating instruments in MIS research. *MIS Quarterly* (13:2), 1989, pp. 146-166.
- [56] Strauss, A. and Corbin, J. *Basics of Qualitative Research*, Sage Publications, Newbury Park, CA, 1990.
- [57] Trocchia, P. J. and Janda, S. How do consumers evaluate Internet retail service quality?. *Journal of Services Marketing* (17:3), 2003, pp. 243-253.

- [58] Wang, R. Y. A product perspective on Total Data Quality Management. *Communications of the ACM* (41:2), 1998, pp. 58-65.
- [59] Yang, Z., Peterson, R. T. and Cai, S. Services quality dimensions of Internet retailing: an exploratory analysis. *Journal of Services Marketing* (17:7), 2003, pp. 685-700.
- [60] Yang, Z., Cai, S., Zhou, Z. and Zhou, N. Development and validation of an instrument to measure user perceived service quality of information presenting web portals. *Information and Management* (42), 2005, pp. 575-589.