SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •
Special Focus

May 2013, Vol. 56 052101:1–052101:14
doi: 10.1007/s11432-013-4835-4

Beautiful, usable, and popular: good experience of interactive products for Chinese users

LIU ShuQing [1], ZHENG XianJun Sam[1,2*], LIU GuanMin[1], JIAN Jiang[1] & PENG KaiPing[1,3*]

[1] *Tsinghua University, Beijing* 100084*, China;*
[2] *Siemens Corporate Research, Princeton, NJ* 08540*, USA;*
[3] *University of California at Berkeley, CA* 94720*, USA*

Received December 5, 2012; accepted January 29, 2013

Abstract User experience (UX) is essential for creating successful interactive products. However, whether the evaluation of user experience is universal across different cultures is questionable. In the past decades, a large number of researches have demonstrated consistently that different cultures play an important role in influencing people's perception and cognition of the world. These cultural differences can also influence people's aesthetic judgments and experience of products. The main goals of our research were two-fold: 1) to develop a standardized Chinese questionnaire for assessing the UX of interactive products, and 2) to evaluate the experience of Chinese users in interacting with three different mobile phones, Nokia, Blackberry, and iPhone. The study results identified three main factors or dimensions contributing to the experience of Chinese users: Hedonic Quality (Stimulation), Pragmatic Quality, and Conformity, as compared to the four dimensions in the AttracDiff, an English questionnaire popularly used in western cultures. Among three mobile products, iPhone was considered the most beautiful, usable, and popular, providing the best user experience overall.

Keywords user experience (UX) design, interactive product, questionnaire, assessment, Chinese, culture and cognition

Citation Liu S Q, Zheng X J S, Liu G M, et al. Beautiful, usable, and popular: good experience of interactive products for Chinese users. Sci China Inf Sci, 2013, 56: 052101(14), doi: 10.1007/s11432-013-4835-4

1 Introduction

User experience (UX) has become more and more essential for creating successful interactive products, as evidenced by numerous Apple's products, such as the iPod, iPhone, and iPad. These successful products seem to share some common characteristics. First, the beautiful design creates the first good impression of the product; and therefore it can increase people's interests in buying and using the product. Secondly, good design also improves the communication between products and users, making the product easy to learn and use. Moreover, a product with good UX not only increases the users' loyalty to the brand, but also attracts more and more new customers.

*Corresponding author (email: sam.zheng@gmail.com, kppeng@berkeley.edu)

User experience is considered to extend beyond usability, where the latter traditionally has been focusing on the functional aspects of a product. Whiteside et al. [1] argued that experience of a product is more important than the efficiency and learnability of a product. Carroll et al. [2] shared similar views, arguing that users are ultimately pursuing a pleasant experience of a product, in stead of focusing on the functioning of the product itself. Indeed, studies by Mahlke [3] showed that the non-functional aspects of the systems, e.g., aesthetics or identity, have significant impact on users' overall assessment of the product. Interacting with attractive products creates more positive emotions for the users than interact with less attractive ones [4]. Furthermore, the products that have been perceived as aesthetically pleasing will also been perceived to have good usability [5].

1.1 What is user experience?

Despite its popular use, the meaning of UX has been ambiguous. There lacks a consensus definition of UX.

Alben's [6] definition of UX is as follows:

"By 'experience' we mean all the aspects of how people use an interactive product: the way it feels in their hands, how well they understand how it works, how they feel about it while they're using it, how well it serves their purposes, and how well it fits into the entire context in which they are using it. If these experiences are successful and engaging, then they are valuable to users and noteworthy to the interaction design awards jury. We call this 'quality of experience'."

Forlizzi et al. [7] defined UX as the results of users interacting with products, which is a dynamic and changing process.

Hassenzahl et al. [8] proposed that UX is:

"A consequence of a user's internal state (predispositions, expectations, needs, motivation, mood, etc.), the characteristics of the designed system (e.g. complexity, purpose, usability, functionality, etc.) and the context (or the environment) within which the interaction occurs (e.g. organisational/social setting, meaningfulness of the activity, voluntariness of use, etc.)"

International Standard Group also attempted to (ISO 9241-210) [9] define UX as:

"All aspects of the user's experience when interacting with the product, service, environment or facility. It includes all aspects of usability and desirability of a product, system or service from the user's perspective"

From above mentioned various definitions, the general consensus is that UX is a multi-dimensional construct that describes users' experience of interacting with a product over a certain period of time. No matter which definition will be agreed and accepted by everyone, how to measure UX is more important from practitioners' point of view. Indeed, a variety of methods have been used to measure and assess the UX of a product. Qualitative methods include user interviews, focus groups, ethnography studies, etc. Quantitative methods include standardized questionnaires, behavioral experiments, and eye tracking methods, etc. And among them, the questionnaire method stands out as a most cost-effective way of accessing the UX of products. It collects quantitative user feedback, and can be administered without environmental and equipment constraints. Therefore, it is often used during the design process to collect users' feedback in a structural way, and is also used for evaluation after the product is developed or released into the market.

1.2 How to measure UX

A number of user questionnaires have been developed, including QUIS (Questionnaire for User interaction Satisfaction) [10], SUMI (Software Usability Measurement Inventory) [11], CSUQ (Computer System Usability Questionnaire) [12] , and SUS (System Usability Scale) [13]. Even though these questionnaires have good validity and reliability, they are heavily focused on the usability, or ease of use of the product. Hassenzahl and colleagues have been developing a standardized questionnaire for assessing people's experience of using a product [14–19]. Hassenzahl et al. [16] first proposed the AttracDiff1, which is composed of 23 items; each item is expressed by a pair of bipolar word anchors (e.g., Confusing-Clear,

Ugly-Beautiful), and is scored with a 7-point Likert scale (e.g. −3 strongly disagree to 3 strongly agree). These 23 items are grouped into three dimensions: ergonomic quality (EQ), hedonic quality (HQ), and appeal. Ergonomic quality refers to the usability measure of a product design and focuses on goal/task-related functions and design, addressing the users' need for utility and usability; the name of ergonomic quality was later changed to pragmatic quality (PQ) [19]. In contrast, hedonic quality measures the originality, novelty, and aesthetics of a product, which are not directly related to users' task goals but address the human needs for change and social identity. Hassenzahl et al. [17] further divided the hedonic quality into two dimensions: stimulation (HQ-S) and identity (HQ-I) as the AttracDiff1 was evolved into the AttracDiff2. Stimulation (HQ-S) relates to personal growth as reflected by the need for stimulation, novelty, and challenges, whereas identity (HQ-I) implies perceived possibility of expressing and communicating one's self or identity to others through using certain products. Based on the perception of PQ and HQ, users may further develop a global judgment of the product's appealingness (APPEAL).

Researchers have used the AttracDiff questionnaire to assess the UX of various products, including MP3 player software [16], the computer display [14], and website design [19]. Good reliability and validity of the questionnaire were shown in all these studies. The questionnaire has so far been widely used in various settings.

1.3 Culture difference or influence

The practice of UX in China has grown considerably in the past several years. More and more Chinese people registered as UPA (Usability Professional Association) members, and more and more companies employ the UX designer, partially inspired by the success of Apple. Questionnaires are often applied to access the user experience of Chinese users in interacting with products. It is, however, still unknown whether the AttracDiff questionnaire can be translated and applied directly to evaluate Chinese users' experience of interactive products. In our opinion, the AttracDiff was developed mainly for Westerners, so it may not be directly applicable to Chinese users.

Indeed, a lot of researches in the past decade [20–23] have demonstrated that cultural difference plays an important role in influencing people's perception and cognition of the world. For instance, the cognitive style of people living in East Asian cultures is more holistic due to their dialectical thinking; whereas people living in Western cultures are more analytic because of their linear thinking. East Asian people tend to view situations as a whole, and are sensitive to the relationships between objects and their context; whereas Western people are inclined to focus on the salient object, isolating it from the background, analyzing its attributes, and trying to classify the object based on some logic.

Moreover, cultural differences also affect the concept of self. The independent self, which is often exhibited by westerners, emphasizes the difference between oneself and others, and the definition of self builds upon individual's behavior, motivations, and value. Whereas the interdependent self, exhibited by easterners, emphasizes the similarity and consistency between oneself and others, and the definition of self is basing on relationship, social identity, and expectations from others. Researches have shown that different self concepts can impact individual's cognition, emotion, and motivations [20,21].

In fact, the cultural differences are also reflected by the different characteristics of the artifacts that East Asians and Westerners create. For instance, the living environments of East Asians are more complex and contain more objects than those of Westerners. Art pieces produced by East Asian artists focused on the contextual information and de-emphasize individual objects, including people [20]. In contrast, Western art work emphasizes individual objects and people with less attention to the field environment [22]. Masuda et al. (2008) [23] conducted several studies to examine cultural differences in painting and photography. For example, Figure 1 illustrates pictures drawn by American and East Asian participants. The left was drawn by an American participant, and the right was drawn by an East Asian participant. The picture drawn by the East Asian contained more objects and had more complex relationships as compared to the one drawn by the American.

Furthermore, Zhang et al. [24] explored the effects of self concept on the aesthetic judgments of angular shapes and rounded shapes. They found people with an independent self, which is often represented by western culture, perceived angular shapes as more attractive; whereas people with the interdependent

Figure 1 Pictures drawn by (a) American and (b) East Asian (adapted from Masuda et al. [23]).

Figure 2 Independent self perceived angular shapes as more attractive; interdependent self found rounded shapes more attractive (adapted from Zhang et al. [24]).

self, which is often represented by East Asian culture, found rounded shapes more attractive. Examples are illustrated in Figure 2.

Overall, the above-mentioned researches have shown that cultural differences between East Asians and Westerners can lead to fundamental differences in their perceptual process and cognitive style, and these culture differences can also influence people's aesthetic judgment and experience.

1.4 Research questions

Our current research project is inspired by the research findings above. First of all, we wanted to explore how Chinese users evaluate their experience of interactive products. Specifically, we wanted to study whether the four dimensions (PQ, HQ-S, HQ-I, and APPEAL) proposed by Hassenzahl and colleagues in AttracDiff are applicable to Chinese users. Throughout this study, we also hoped to develop a standardized Chinese questionnaire for assessing the UX of interactive products, i.e., the Chinese UX Questionnaire. Another goal of the study was to explore differences in the experience of Chinese users in interacting with various cell phones, such as Nokia, Blackberry, and iPhone.Based on the current market reviews, we hypothesized that Nokia phone is practical, and Blackberry is powerful, and iPhone is a good combination of design and functionality. According to Hassenzahl (2000), users' task difficulty rating can be used as a good validity measure against with their UX rating of a product. For instance, if users spend more mental effort in completing a task, namely the more energy an individual has to activate to meet the perceived task demand, then the product is less likely to be usable and quite likely will have a low Ergonomic Quality score. Also the more mental effort it requires, the less likely users will like to use the product, and therefore the level of the mental effort can be associated with the level of appealing (Appeal). Lastly, because the Hedonic Quality is not related to the tasks, there should be no relationship between the task difficulty rating and the HQ rating.

In order to test the hypothesis that cultural differences influence our experience of products, we asked the subjects to rate the products' UX and the degree of difficulty after they finished the requested tasks. Firstly, we conducted a pilot study, which is composed of a semi-structured interview. We showed the pictures of cell phones and laptops, asked the subjects to describe their feelings when using the products and what ideal interactive products should be like. Based on AttracDiff2 and the results of our interviews, we developed the initial Chinese edition of UX questionnaire. Secondly, we asked the subjects to rate the degree of difficulty after completing the requested tasks with the three kinds of cell phones, Nokia, Blackberry, and iPhone, and then completed the UX questionnaire. Lastly, we conducted an item analysis,

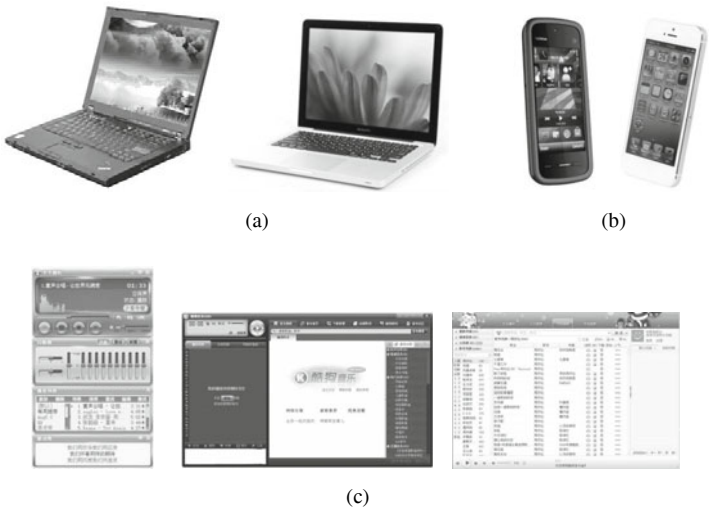

Figure 3 Various interactive products used in the pilot study. (a) Laptops (ThinkPad & Macbook); (b) cell phones and laptops (Nokia & iPhone); (c) MP3 music-player.

reliability and validity analysis on the UX questionnaire. We compared the difficulty degrees and UX ratings of each cell phone and conducted a correlation analysis between them.

2 Method

2.1 Pilot study

In order to confirm whether the four identified dimensions (PQ, HQ-S, HQ-I, and APPEAL) and their associated specific items in AttracDiff are suitable to measure Chinese users' experience of interactive products, and also to explore if there are other dimensions and items for Chinese users, we conducted a semi-structured interview to elicit users' impressions and ratings of their product experience with 26 college students (12 females, 14 males; ages 19–26).

First, each of the interviewees was individually presented with pictures of two brands of cell phones and laptops (Nokia & iPhone; ThinkPad & Macbook, see Figure 3 (a) and (b)). Then they were asked to list the most important aspects according to their own impression and experience (at least 3 adjective for each aspects). Secondly, they were separately required to use three different types of MP3 music-player software (see Figure 3(c)) and finish five tasks using the software. After the completion of tasks, they were prompted to talk about their feelings from various perspectives, such as pragmatic quality, hedonic quality, attractiveness, identity, and appeal, and they were encouraged to come up with any additional evaluation dimensions about products. The interview contents were recorded and coded.

The analysis of coded results showed that all the interviewees expressed their general impression of interactive products from all four aspects as used in AttracDiff, i.e, Pragmatic Quality, Hedonic Quality (Stimulation, Identity), Appeal. In addition, a number of participants introduced another evaluation dimension, which we called Conformity, including items of how popular or market prevalence a product is, whether there are already many users, and what their friends think about the product, etc. The results are seen in Table 1.

The result of the pilot study is a preliminary version of the Chinese UX questionnaire with a total of 52 items, a collection of all items used in AttrakDiff2, and the new items received from the interviews. The initial Chinese UX questionnaire comprises five dimensions: Pragmatic Quality, Stimulation (Hedonic Quality), Identity (Stimulation), Appeal, and Conformity. Each item is expressed as a pair of bipolar Chinese idioms and expressions, for example, Pragmatic Quality: 简单明了 (simple) -错综复杂 (complex), Stimulation: 新颖独特 (Novel)-平凡普通 (Ordinary), Identity: 新潮时尚 (Stylish)-陈旧俗

Table 1 Number of interviewees mentioning various dimensions

Dimension	Numbers	Percentages (%)	Examples
Pragmatic quality	26	100	usable, reliable, easy to carry, fast
Stimulation	12	46	novel, relax, nice color, innovative, flesh, dynamic
Identity	4	15	showing off, different
Appeal	8	31	self-satisfactory, trendy, flashy, simple, attractive
Conformity	9	35	popular, good remarks by others

(a) (b) (c)

Figure 4 Three types of cell phones used in the study. (a) Nokia 6303; (b) Blackberry 8900; (c) iPhone 4.

气 (Tacky), Appeal: 令人排斥 (Rejecting) -充满魅力 (Inviting), Conformity: 闻名遐迩 (Famous) -默默无闻 (Unknown). Each item is scored using the same 7-point Likert scale as in AttracDiff.

2.2 Main study

The main purpose of the study was to develop a Chinese UX Questionnaire with clear structure and adequate reliability and validity; one that will accurately assess the Chinese users' experience of interactive products.

2.2.1 *Participants*

Seventy two college students at Tsinghua Chinese university (36 males and 36 females, with normal visual acuity) were recruited to participate in the study, receiving monetary compensation of RMB 20. The sample's mean age was 23 years (ranging from 18 to 27).

2.2.2 *Cell phone types*

We chose to use cell phones in this study due to their popular usage and their importance in our daily lives and experience. The three tested cell phone types were (a) Nokia 6303 (silver); (b) Blackberry 8900 (black); and (c) iPhone 4 (white) as illustrated in Figure 4.

2.2.3 *Measures*

The After-Scenario Questionnaire and the Chinese UX Questionnaire were used in our study.

The After-Scenario Questionnaire [25] was applied to rating the task Difficulty degree in a Likert format. It is a uni-dimensional rating scale, and consists of 9 seven-point scale items from 1 for "very easy" to 7 for "very difficult". The Chinese UX Questionnaire consists of 52 seven-point scale items with opposite anchors of Chinese idioms and expressions. The scale was used to measure Pragmatic Quality, Stimulation (HQ), Identity (HQ), Appeal, and Conformity of the cell phones.

2.2.4 *Procedure*

The experiment was conducted in the laboratory of the Psychology Department at Tsinghua University. Each participant was led separately into the laboratory. After a brief introduction and short instructions by the experimenter, each participant performed nine different scenario tasks with each cell phone. The

Table 2 The task scenarios used in the main study

No.	Description of task scenarios
Task 1	You find the cell-phone number of your friend, and then send a message to her and ask her whether you can participate the psychology experiment. (Completion criteria: SMS Success.)
Task 2	You read the new message from your friend and know the experiment arrangement. (Completion criteria: Read the message.)
Task 3	You check your schedule and find you can attend the experiment, so you reply OK. (Completion criteria: Find no class and no other personal engagement during the experimental time.)
Task 4	You write down the experiment time and location in the calendar as a reminder. (Completion criteria: Input relevant information in the calendar.)
Task 5	Suppose it is the evening of May 18. You set your alarm clock for 7 o'clock tomorrow morning. (Completion criteria: Finish setting clock.)
Task 6	It is 8:30 am on May 19. You set out and listen to music on the bus. (Completion criteria: Find music in the cell-phone, and then listen for 1 minute.)
Task 7	You arrive at the psychology department, but you can't find the experiment lab. So you call your friend to ask the location of the room. (Completion criteria: You are informed with the lab's position.)
Task 8	The layout of the experimental lab looks interesting. So you take a picture. (Completion criteria: Take a picture.)
Task 9	You try out some other phone functionalities, e.g, games, while waiting for the beginning of the experiment. (Completion criteria: Experience freely for 2 minutes.)

background of the scenario tasks was that: "Imagine you are a freshman student, and you are interested in participating in a psychology experiment. You want to contact your friend at the psychology department to set up an appointment for the study."

The task scenarios as listed in Table 2 included three clusters of functions or tasks: Basic functions—receiving messages, sending messages, making calls; Auxiliary functions—setting alarms and schedules; Entertaining functions—listening to music, taking pictures, playing games. Immediately after each scenario task, the difficulty degree was assessed with the ASQ.

After having rated all scenario tasks, the participants were asked to fill out the Chinese UX Questionnaire. All questionnaires were presented in paper-pencil format. This procedure was repeated for every cell phone type. Note that the order of the cell phone was counter-balanced to offset fatigue and learning effects. The whole experiment lasted for about 1 h.

3 Results

3.1 Item-total correlations

In order to delete the items that do not measure the same qualities as the whole scale, the original 52 items were analyzed by correlation analysis to establish acceptable item-total correlations. Those that had a low correlation mean that they don't measure the same qualities as the whole scale. The criteria for item reduction were as follows. First, the item-total correlations were significant ($p < 0.05$). Second, the items with the item-total correlations coefficients below 0.4 were eliminated. The two criteria above guarantee the items measure the same qualities as the whole scale. Consequently, we deleted ten items not up to the criteria.

3.2 Exploratory factor analysis

The remaining 42 items were subjected to the exploratory factor analysis (i.e., principle components with varimax rotation); the results show KMO = 0.953 and Bartlett's Test $\chi^2 = 9167.898$, $p < 0.001$. The level is significant, indicating that the data are suitable for factor analysis. Six factors with eigenvalue greater than 1.0 emerged; however, three of them consisted of less than 3 items. Another several rounds of

Table 3 The remaining 20 items and their loadings in three identified principal components/factors after the exploratory factor analysis

	Factor 1	Factor 2	Factor 3
T33 平凡普通 (Ordinary)-新颖独特 (Novel)	0.871		
T32 低档廉价 (Cheap)-高档昂贵 (Premium)	0.896		
T36 陈旧俗气 (Obsolete)-新潮时尚 (Stylish)	0.858		
T29 缺乏想象 (Unimaginative)-富有创造 (Creative)	0.820		
T34 中规中矩 (Conventional)-独出心裁 (Inventive)	0.851		
T16 功能单调 (Weak)-功能强大 (Powerful)	0.785		
T26 传统保守 (Conservative)-改革创新 (Innovative)	0.769		
T23 乏味消极 (Discouraging)-生动积极 (Motivating)	0.781		
T22 手工操作 (Manual manipulation)-智能控制 (Intelligent control)	0.793		
T20 更新缓慢 (Slowly updated)-更新迅速 (Quickly updated)	0.644		
T2 大众化 (Working-class)-贵族化 (Aristocratic)	0.737		
T43 常见普通 (Common)-与众不同 (Distinctive)	0.783		
T49 小心谨慎 (Cautious)-大胆豪放 (Bold)	0.679		
T41 难以操作 (Unruly)-易于操作 (Manageable)		0.834	
T38 让人怀疑 (Doubtful)-值得信赖 (Trustworthy)		0.774	
T39 繁琐累赘 (Cumbersome)-灵活高效 (Straightforward)		0.730	
T44 隔离我与他人的关系 (Separates me from people)-拉近我同他人的距离 (Brings me closer to people)		0.645	
T10 娱乐化 (Entertaining)-商务化 (Commercial)			0.751
T40 用户甚少 (Rarely used)-用户众多 (Widely used)			0.786
T6 小众流行 (Smallspread)-广泛流行 (Widespread)			0.786

Table 4 Results of confirmatory factor analysis

Index	χ^2	df	χ^2/df	NFI	RFI	IFI	CFI	RMSEA
Value	236.326	148	1.597	0.877	0.842	0.950	0.949	0.075

factor analysis were conducted for item reduction with the following criteria. First, each factor consisted of no less than 3 items, because measurement of less than 3 items is not stable. Second, the loading of each item on the main factor should be above 0.4, which means they measure the same quality as other items do. Third, items with loading above 0.3 on two or more factors were deleted, which means items measuring two or more factors are not included. Finally, 20 items were attained as shown in Table 1. The results show KMO = 0.906 and Bartlett's Test $\chi^2 = 1621.414$, $p < 0.001$.

As illustrated in Table 3, three main principal components or factors were identified after the exploratory factor analysis. Factor 1 was similar to the original Simulation (HQ) dimension, Factor 2 was similar to the original Pragmatic Quality dimension, and Factor 3 consisted of 3 items on the Conformity dimension. Therefore, the resulting 20 items constituted three dimensions: Simulation (HQ), Pragmatic Quality and Conformity. Three factors can account for 69.237% total variance.

3.3 Confirmatory factor analysis

The hypothesized model of the scale was estimated via maximum likelihood (ML) with software AMOS, and the results were as follows. The CFI of the model was 0.949, and the RMSEA of the model was 0.075. According to previous literature, the CFI values of at least 0.90 are considered adequate for good models [17] and the RMSEA values below 0.08 reflect an acceptable error of approximation. Therefore, the hypothesized model was fitted to the data. The results are illustrated in Figure 5 and Table 4.

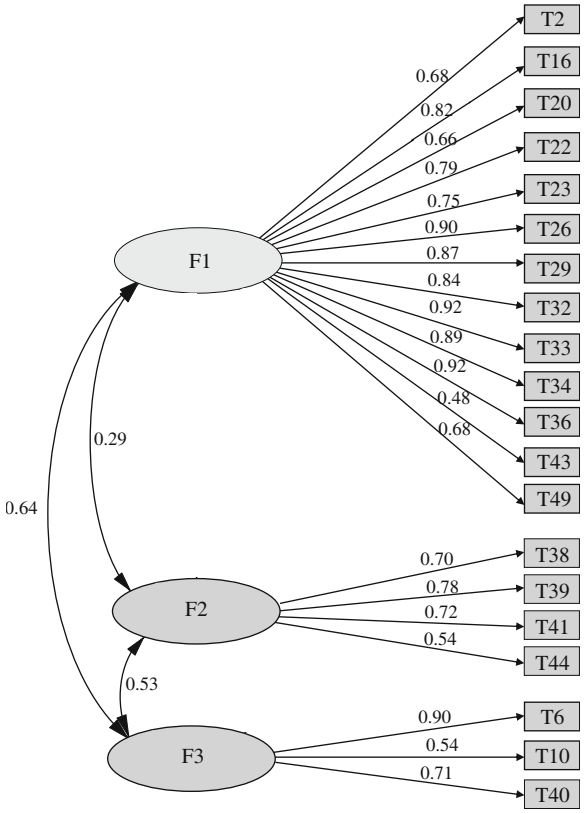

Figure 5 The result from the confirmatory factors analysis has shown three factors and their associated 20 items.

Table 5 Internal consistency results (based on the first half of data)

	Stimulation	Pragmatic quality	Conformity	Overall
Cronbach's α	0.956	0.772	0.782	0.931

Table 6 Internal consistency results (based on the second half of data)

	Stimulation	Pragmatic quality	Conformity	Overall
Cronbach's α	0.960	0.769	0.737	0.939

3.4 Internal consistency

We also evaluated the internal consistency of the resultant 29-items scale using Cronbach's α. The higher the coefficient, the higher the internal consistency of the items, which mean they measure the same quality. Cronbach's α of the whole scale, the first factor (20 items), the second factor (6 items) and the third factor (3 items) were between 0.85 and 1 (Tables 5 and 6), and Cronbach's α of the total scale was 0.951. All of them are higher than 0.70, showing good internal consistency for the whole questionnaire as well as its three factors.

3.5 Tasks difficulty rating

The results of task difficulty rating for three phones on various tasks are illustrated in Figure 6. The detailed statistical analyses are as follows.

3.5.1 *The general tasks' difficulty degree*

We analyzed data by use of repeated analysis of variance (ANOVA). The cell phone type was the independent variable and the difficulty degree of overall function task was the dependent variable. The main

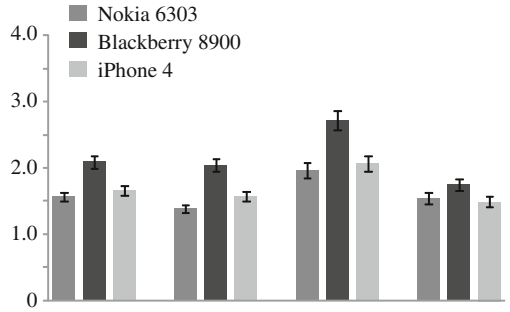

Figure 6 Tasks difficulty ratings of three cell phones.

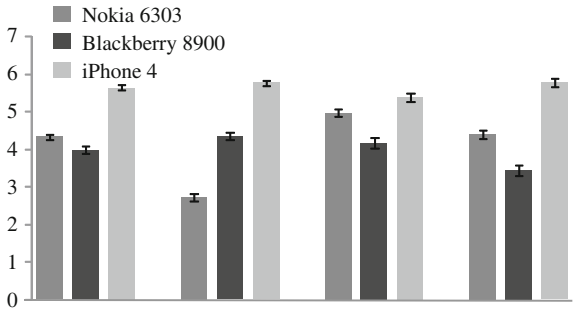

Figure 7 UX ratings of three cell phones.

effect of cell phone type was significant, $F(2, 142) = 26.627$, $p < 0.001$. The results of LSD Multiple Comparisons among different types of cell phones showed that with regard to the overall function tasks, using Nokia 6303 was significantly easier than Blackberry 8900, $p < 0.01$. There was a significant difference between iPhone 4 and Blackberry 8900, $p < 0.001$. There was no difference in overall tasks' difficulty degree between Nokia 6303 and iPhone 4. Therefore, there were significant differences in general tasks' difficulty degree between Blackberry 8900 and Nokia 6303/iPhone 4.

3.5.2 *The basic function tasks' difficulty degree*

A repeated analysis of variance (ANOVA), with cell phone type as the independent variable and the basic functions task difficulty degree as the dependent variable revealed that the main effect of cell phone type was significant, $F(2, 142) = 30.055$, $p < 0.001$. The result of LSD Multiple Comparisons among different types of cell phones showed that with regard to the basic functions tasks, using Nokia 6303 was significantly easier than iPhone 4 ($p < 0.05$) and Blackberry 8900 ($p < 0.001$). There was a significant difference in basic function tasks' difficulty degree between Blackberry 8900 and iPhone 4 , $p < 0.001$. Therefore, there were significant differences in the basic functions task difficulty degree among the three cell phones.

3.5.3 *The auxiliary function tasks' difficulty degree*

A repeated analysis of variance (ANOVA), with cell phone type as independent variables and the auxiliary functions task difficulty degree as dependent variable revealed that the main effect of cell phone type was significant, $F(2, 142) = 14.869$, $p < 0.001$. The results of LSD Multiple Comparisons among different types of cell phones showed that with regard to the auxiliary functions tasks, using Nokia 6303 was significantly easier than Blackberry 8900, $p < 0.001$, and using iPhone 4 was significantly easier than Blackberry 8900, $p < 0.001$. There was no difference in auxiliary function tasks' difficulty degree between Nokia 6303 and iPhone 4, $p > 0.05$. Therefore, there were significant differences in auxiliary function tasks' difficulty degree between Blackberry 8900 and Nokia 6303/iPhone 4.

3.5.4 *The entertaining function tasks' difficulty degree*

A repeated analysis of variance (ANOVA), with cell phone type as the independent variable and the entertaining functions task difficulty degree as dependent variable revealed that the main effect of cell phone type was significant, $F(2, 142) = 14.869$, $p < 0.05$. The results of LSD Multiple Comparisons among different types of cell phones showed that with regard to the entertaining functions tasks, using Nokia 6303 was significantly easier than Blackberry 8900, $p <= 0.05$, and using iPhone 4 was significantly easier than Blackberry 8900, $p < 0.05$. There was no difference in auxiliary function tasks' difficulty degree between Nokia 6303 and iPhone 4, $p > 0.05$. Therefore, there were significant differences in entertaining functions task difficulty degree between Blackberry 8900 and Nokia 6303/iPhone 4.

3.6 UX rating

The results of UX rating for three phones on various dimensions are illustrated in Figure 7. The detailed statistical analyses are as follows.

3.6.1 *The general UX rating differences*

We analyzed data by use of repeated analysis of variance (ANOVA). Cell phone type was the independent variable and the general UX rating score was the dependent variable. The main effect of cell phone type was significant, $F(2, 142) = 152.421$, $p < 0.001$. The result of LSD Multiple Comparisons among different types of cell phones showed that the general UX appraisals of iPhone 4 was significantly higher than Blackberry 8900 ($p < 0.001$) and Nokia 6303 ($p < 0.001$), and the appraisals of Blackberry 8900 were not significantly higher than Nokia 6303, $p > 0.05$. Therefore, there were significantly general UX appraisals among these three phones. Therefore, there were significant differences in general UX appraisals between iPhone 4 and Nokia 6303/Blackberry 8900.

3.6.2 *The pragmatic quality differences*

A repeated analysis of variance (ANOVA), with cell phone type as the independent variable and the Pragmatic Quality rating as the dependent variable revealed that the main effect of cell phone type was significant, $F(2, 142) = 31.099$, $p < 0.001$. The result of LSD Multiple Comparisons among different types of cell phones showed the Pragmatic Quality appraisals of Blackberry 8900 was significantly lower than iPhone 4 ($p < 0.001$) and Nokia 6303 ($p < 0.001$), and the Pragmatic Quality appraisals of Nokia 6303 was significantly higher than Blackberry 8900. Therefore, there were significant differences in Pragmatic Quality appraisals among the three cell phones.

3.6.3 *The stimulation (HQ) differences*

A repeated analysis of variance (ANOVA), with cell phone type as the independent variable and the Stimulation (HQ) as dependent variable revealed that the main effect of cell phone type was significant, $F(2, 142) = 320.678$, $p < 0.001$. The result of LSD Multiple Comparisons among different types of cell phones showed the Stimulation Appraisals of iPhone 4 was significantly higher than Blackberry 8900 ($p < 0.001$) and Nokia 6303 ($p < 0.001$), and the Stimulation appraisals of Blackberry 8900 was significantly higher than Nokia 6303, $p < 0.001$. Therefore, there were significant differences in Stimulation appraisals among the three cell phones.

3.6.4 *The conformity differences*

A repeated analysis of variance (ANOVA), with cell phone type as the independent variable and the Conformity rating as dependent variable revealed that the main effect of cell phone type was significant, $F(2, 142) = 88.255$, $p < 0.001$. The result of LSD Multiple Comparisons among different types of cell phones showed the Conformity appraisal of iPhone 4 was significantly higher than that of Blackberry 8900 and Nokia 6303, and that of Nokia 6303 was significantly higher than that of Blackberry 8900, $p < 0.001$. Therefore, there were significant differences in conformity appraisals among the three cell phones.

3.7 The correlations between task difficulty rating and UX rating

Table 7 shows the correlation results between the task difficulty rating and the UX rating. All the task difficulty ratings were negatively correlated with the General UX Rating, with the General Task's Difficulty ($r = -0.275$, $p < 0.01$), the Basic Function Tasks' Difficulty ($r = -0.234$, $p < 0.01$), Auxiliary Function Tasks' Difficulty ($r = -0.269$, $p < 0.01$) and Entertaining Function Tasks' Difficulty ($r = -0.202$, $p < 0.01$). There were no relationship between the correlation of any task different rating and the Stimulation dimension. The Pragmatic Quality rating correlated negatively with the General Task's Difficulty ($r = -0.446$, $p < 0.01$), the Basic Function Tasks' Difficulty ($r = -0.430$, $p < 0.01$), the Auxiliary Function Tasks' Difficulty ($r = -0.463$, $p < 0.01$) and the Entertaining Function Tasks'

Table 7 Correlations between task difficulty rating and UX rating (**: $p < 0.01$)

	General Tasks' Difficulty	Basic Function Tasks' Difficulty	Auxiliary Function Tasks' Difficulty	Entertaining Function Tasks' Difficulty
General UX Rating	−0.275**	−0.234**	−0.269**	−0.202**
Stimulation	−0.007	0.039	−0.013	−0.058
Pragmatic Quality	−0.446**	−0.430**	−0.463**	−0.233**
Conformity	−0.241**	−0.213**	−0.209**	−0.195**

Difficulty ($r = -0.233$, $p < 0.01$). The Conformity rating also correlated negatively with the General Task's Difficulty ($r = -0.241$, $p < 0.01$), the Basic Function Tasks' Difficulty ($r = -0.213$, $p < 0.01$), the Auxiliary Function Tasks' Difficulty ($r = -0.209$, $p < 0.01$) and the Entertaining Function Tasks' Difficulty ($r = -0.195$, $p < 0.01$). These analysis results indicated that if users exercised more mental effort when interacting with the cell phone, they would feel the cell phone less pragmatic (less usable) and less conformity (less popular), therefore, the lower level of user experience of the cell phone. However, the results also indicated that the Stimulation rating was independent of the mental effort as shown by task difficult rating, which is consistent with Hassenzahl (2000).

4 Discussion

The main purpose of this study was to develop a Chinese UX questionnaire. According to the research findings in culture differences, we were interested in exploring how Chinese users evaluated user experience of interactive products, specifically whether Chinese users rated the interactive products from the four dimensions (PQ, HQ-S, HQ-I, and APPEAL) proposed by Hassenzahl and colleagues [14–19] in AttracDiff.

The study results demonstrated that Chinese users assessed their experience of interactive products from three dimensions: Hedonic Quality (Stimulation), Pragmatic Quality, and Conformity instead of the four dimensions proposed by Hassenzahl and colleagues. This means that a perfect interactive product should be beautiful, usable, and popular. The research findings confirm our hypothesis that culture can also affect users' experience in interacting with products [16–22].

First, the Stimulation dimension of the Chinese version of the UX questionnaire included some items from the Stimulation, Appeal and Identity dimensions in AttracDiff. This showed that Chinese users' UX towards the interactive products are based on the global impression mainly from the Stimulation dimension, rather than more detailed sub-dimensions found in the West. The results provided evidence of the division of East Asians' dialectical thinking style and Westerners' linear thinking style. The former focuses on the relationship of the objects and the background, and the latter on the disposition and categorization of objects [26,27]. Secondly, two factors in AttracDiff (Programmatic Quality and Stimulation) apply to both Western users as well as Chinese users, which suggests that the need for utility and usability, and the need for personal development and growth are universal [15]. However, the other Hedonic Quality-Identity in the AttracDiff, such as the Identity dimension, was not applicable to Chinese users. This can be explained by the different cultures in the East and West. The Identity dimension means people have the need to express themselves through their own objects. The need is stronger for users in Western culture since the culture emphasizes the role of each individual and one's own identity; whereas, the cultures in East Asia, such as China, de-emphasizes the role of each individual; therefore, the need for identity in particular via objects is less strong.

This study finds another important new dimension, Conformity. It was introduced in the Chinese UX questionnaire, which again reflects cultural influence. Westerners make judgments typically based on their own opinions; however, East Asians care more about others' opinions [20,21]. This phenomenon was also confirmed by several manifestations received from the interview, "If my friends like this product, I would like it more," "If many others already use it, I would like to give a try too". These findings are consistent with the previous research on culture and self-concept [20,21]. Thus Western people show a

more independent self vs. East Asians with a more interdependent self.

Similar to Hassenzahl (2000), in our study, we also used users' task difficulty rating as a validity metric against with the UX rating of a product. Hassenzahl argued that users' mental effort as illustrated by task difficulty rating should be correlated with the Pragmatic Quality rating. And our study results confirmed that the more difficult to complete a task in using a product, the lower ratings score of PQ will be given by the users. Even though Hassenzahl also argued that the mental effort can be associated with the rating of Appeal, this dimension was not included in our Chinese UX questionnaire from the exploratory data analysis. Lastly, Hassenzahl reasoned that because the Hedonic Quality is not associated with the tasks, there should be no relationship between the task difficulty rating and the HQ rating. This was consistent with our study results that the task difficult rating was not correlated with the HQ Stimulation rating. What's interesting is that we have introduced a new dimension, Conformity, in the questionnaire, and its rating has negative correlations with the task difficulty. This suggests that if a product requires a high mental effort, it would limit its likelihood to be spread among Chinese users.

Another goal of this study was to systematically explore the experience of Chinese users in interacting with various cell phone products, such as Nokia, Blackberry, and iPhone. Even though marketing reviews or individual critics often praised Nokia phone for its usability, Blackberry phone for its functionality, and iPhone for its richness of experience, there has not been empirical studies to evaluate Chinese users' experience of these three phones. The rating results of the Chinese version of the UX questionnaire showed that the general UX rating of iPhone is the highest. Moreover, iPhone is also rated the highest in the dimensions of Stimulation and Conformity, while Nokia and iPhone are both rated higher than Blackberry in Pragmatic Quality. In terms of task difficulty rating, both iPhone and Nokia are considered easier to operate than Blackberry. These study findings further confirm the notion that good UX: beautiful, usable, and popular, is critical to the success of interactive products in Chinese market. Therefore, it is not surprising that iPhone is dominating the current Chinese cell phone market (even though another product with good user experience, Samsung Galaxy, is also becoming successful in China). Nokia once was flourishing in China and the world due to its high pragmatic quality. However, pragmatic quality alone doesn't produce rich user experience. As to Blackberry, its complexity and poor UX rating probably can account partially its failure in China. These findings can also provide implications for Chinese phone manufacturers.

One of limitations of the current study is that we only collected data from 72 Chinese college students, even though similar issues also existed in previous studies [e.g., 14,17]. The findings should be generalized to the general public with caution. Future studies should include more users from different backgrounds and age groups to increase the questionnaire's validity.

5 Conclusions

In sum, we developed a Chinese questionnaire to assess the users' experience of the design of interactive products. The initial Chinese UX Questionnaire was based on AttracDiff, a widely used English questionnaire, which was proposed and refined by Hassenzahl and his colleagues. It originally extended with an additional dimension and several new items received from the pilot study of Chinese users.

We also conducted a study of applying the initial questionnaire to assess the users' experience of interacting with three different cell phones, Nokia, Blackberry, and iPhone. Seventy-two Chinese participants were involved. The result of both exploratory factor analysis and confirmatory factor analysis showed that three factors or dimensions were identified: Hedonic Quality (Stimulation), Pragmatic Quality, and Conformity, whereas there were four dimensions in the AttracDiff2. These findings demonstrated that cultures not only influence people's cognition, but also their experience of interacting with products.

More user data should be collected to refine the questionnaire and the questionnaire should be used to assess other interactive products in the future. Preferences among different genders, districts, and social class are also worth exploring.

Acknowledgements

We thank Wang RuiQi and Feng YuHan for data collection, and three anonymous reviewers for providing helpful comments. We also thank Tim Beneke's proof reading of the paper.

References

1 Chang H C, Lai H H, Chang Y M. A measurement scale for evaluating the attractiveness of a passenger car form aimed at young consumers. Int J Ind Ergon, 2007, 37: 21–30
2 Carroll J M, Thomas J C. Fun. ACM SIGCHI Bull, 1988, 19: 21–24
3 Mahlke S. Aesthetic and symbolic qualities as antecedents of overall judgements of interactive products. In: Bryan-Kinns N, Blanford A, Cruzon P, et al., eds. People and Computers XX—Engage. Heidelberg: Springer, 2007. 57–64
4 Sonderegger A, Sauer J. The influence of prototype fidelity and aesthetics of design in usability tests: Effects on user behaviour, subjective evaluation and emotion. Appl Ergon, 2010, 41: 403–410
5 Seva R, Gosiaco T, Santos D. Product design enhancement using apparent usability and affective quality. Appl Ergon, 2011, 42: 511–517
6 Alben L. Quality of experience: Defining the criteria for effective interaction design. Interactions, 1996, 3: 11–15
7 Forlizzi J, Battarbee K. Understanding experience in interactive systems. In: Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques. New York: ACM, 2004. 261–268
8 Hassenzahl M. Tractinsky N. User experience-a research agenda. Behav Inf Technol, 2006, 25: 91–97
9 Bevan N. UX, usability and ISO standards. http://www.cs.tut.fi/ihte/CHI08_workshop/slides/Bevan_UXEM_slides.pdf
10 Chin J P, Diehl V A, Norman K L. Development of an instrument measuring user satisfaction of the human-computer interface. In: Proceedings of the Conference on Human Factors in Computing Systems. New York: Association for Computing Machinery, 1988. 213–218
11 Kirakowski J, Corbett M. SUMI: The software usability measurement inventory. Br J Educ Technol, 1993, 24: 210–212
12 Lewis J R. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for user. Int J Hum-Comput Interact, 1995, 7: 57–78
13 Brooke J. SUS: A "quick and dirty" usability scale. In: Jordan P W, Thomas B, McClell I L, et al., eds. Usability Evaluation in Industry. London: Taylor and Francis, 1996. 189–194
14 Hassenzahl M. The effect of perceived hedonic quality on product appealingness. Int J Hum-Comput Interact, 2001, 13: 481–499
15 Hassenzahl M. The interplay of beauty, goodness, and usability in interactive products. Hum-Comput Interact, 2004, 19: 319–349
16 Hassenzahl M, Platz A, Burmester M, et al. Hedonic and ergonomic quality aspects determine a software's appeal. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York: ACM, 2000. 201–208
17 Hassenzahl M, Burmester M, Koller F. AttrakDiff: Ein Fragebogen Zur Messung wahrgenommener hedonischer und pragmatischer Qualitt [AttracDiff: A questionnaire to measure perceived hedonic and pragmatic quality]. In: Ziegler J, Szwillus G, eds. Mensch & Computer 2003. Interaktion in Bewegung. Stuttgart: B. G. Teubner, 2003. 187–196
18 Hassenzahl M. The effect of perceived hedonic quality on product appealingness. Int J Hum-Comput Interact, 2002, 13: 479–497
19 Hassenzahl M, Kekez R, Burmester M. The importance of a software's pragmatic quality depends on usage modes. In: Proceedings of the 6th International Conference on Work With Display Units. Berlin: ERGONOMIC Institut für Arbeits-und Sozialforschung, 2002. 275–276
20 Kitayama S, Duffy S, Uchida Y. Self as cultural mode of being. In: Handbook of Cultural Psychology. New York: Guilford Press, 2007. 136–174
21 Markus H R, Kitayama S. Culture and the self: Implications for cognition, emotion, and motivation. Psychol Rev, 1991, 98: 224–253
22 Nisbett R E, Peng K P, Choi I, et al. Culture and systems of thought: Holistic versus analytic cognition. Psychol Rev, 2001, 108: 291–310
23 Masuda T, Gonzalez R, Kwan L, et al. Culture and aesthetic preference: Comparing the attention to context of East Asians and Americans. Pers Soc Psychol Bull, 2008, 34: 1260–1275
24 Zhang Y, Feick L, Price L J. The impact of self-construal on aesthetic preference for angular versus rounded shapes. Pers Soc Psychol Bull, 2006, 32: 794–805
25 Lewis J R. Psychometric evaluation of an afterscenario questionnaire for computer usability studies: The ASQ. ACM SIGCHI Bull, 1991, 23: 78–81
26 Nisbett R E, Masuda T. Culture and point of view. Proc Natl Acad Sci U S A, 2003, 100: 11163–11170
27 Nisbett R E, Miyamoto Y. The influence of culture: holistic versus analytic perception. Trends Cogn Sci, 2005, 9: 467–473