

## Article

# An Empirical Study on User Experience Evaluation and Identification of Critical UX Issues

Lin Feng \* and Wei Wei

Department of Economics, School of Economics and Finance, Xi'an Jiaotong University, Xi'an 710049, China; wei\_wei@mail.xjtu.edu.cn

\* Correspondence: fenglin@migu.cn

Received: 5 March 2019; Accepted: 19 April 2019; Published: 24 April 2019



**Abstract:** We introduce an approach that supports researchers and practitioners to determine the quality of first-time user experience (FTUX) and long-term user experience (LTUX), as well as to identify critical issues with these two types of UX. The product we chose to study is a mobile fitness application. Mobile apps tend to have a much shorter service life than most other products; thus, the developers/designers need to pay great attention to both first-time and long-term user experience. This study is based on a multi-method approach. We employed the AttrakDiff questionnaire to assess users' first impressions of the app, and the UX Curve method to evaluate how users' experience of the app has changed over time. Besides the quantitative data, which helped to determine the quality of user experience, we also collected qualitative data during two interviews with participants, and focused on the issues that predominantly deteriorated user experience. A four-coordinate plane tool was designed later in the data analysis process that combined the two kinds of user experience data at the same time, which led to a qualitative positioning of the user experience status of a certain product. The model was further successfully adopted in the identification of user experience issues of an online fitness application.

**Keywords:** user experience evaluation; AttrakDiff questionnaire; the UX Curve method; sustainability; mobile application development

## 1. Introduction

Early in 1998, Pin and Gilmore claimed that “from now on, leading-edge companies—whether they sell to consumers or businesses—will find that the next competitive battleground lies in staging experiences” [1]. Experiences, as they suggested, have emerged as “the progression of economic value” [1]. Good user experience (UX) is desirable for both users and companies. For users, a delightful interaction experience not only fulfils their instrumental needs, but also non-instrumental ones such as beauty, satisfaction, and pleasure [2,3], while for companies, a delightful user experience is regarded as one of the key drivers of sustainability development [4,5]. Evidence shows that providing positive UX can increase user satisfaction and loyalty, thus promoting the commercial success of the company [6]. For products that are unable to offer good user experience, UX measurement feedback can assist development teams to fix their experiential problems, therefore improving the UX quality of the product [7,8].

“User experience” is a concept that includes all aspects of how individuals interact with a product [9,10]. The concept of UX is highly context-dependent and dynamic [7,11]. We respect its multipipeline nature, while, in this study, we adopt Hassenzahl and Tractinsky's definition of UX: “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., organizational/social setting,

meaningfulness of the activity, voluntariness of use, etc.)” [2]. Since the 1990s, UX has become a popular tool to evaluate human beings’ interactions with a product, system, or service [12–14]. In contrast to usability research, which predominantly focused on task efficiency, UX research has emphasized experiential qualities. There are a few fundamental UX principles:

1. UX takes a holistic perspective of the user–product interaction [15], including use of the products, as well as the meaning and emotion among users through their interactions with products [16].
2. UX focuses on both pragmatic values and hedonic value [6]. In general, UX studies explore the relationships between usability, symbolic, and aesthetic value for user experience with products.
3. UX emphasizes the importance of the context of use [2], as different usage contexts can result in different experiences.

Both qualitative and quantitative approaches can be applied in UX research. Common qualitative approaches include interviews, focus groups, and user observation [17]. UX metrics provide quantitative data, which can help to measure and track user experience. Popular UX metrics that are being used in studies are the Usability Scale (SUS), Customer Satisfaction Score (CSAT), Net Promoter Score (NPS), and Conversion rate. We listed only a very few here; for UX researchers, they can choose many other UX metrics depends on their goals.

Mobile apps have become more and more common in people’s daily lives. We have observed in the market that a task-oriented mobile application that fails to satisfy people’s non-instrumental needs, such as enjoyability and pleasure, is still likely to result in a market failure, and many apps that offer a lot of powerful features still cannot win popularity among users. We believe that the UX level of the products is the main reason behind this phenomenon. On one hand, when there are a large number of apps offering similar functions, a great first impression of the product experience often leads to users’ decision of which application to choose from. On the other hand, a pleasant long-term experience is the key to users’ loyalty and continuity towards a certain application. Thus, the first-time experience (FTUX) and long-term experience (LTUX) both play vital roles in mobile apps users’ life cycle, and need to be paid equal attention when studied.

Most UX studies regarding to mobile apps have either focused on users’ short-term experience, or long-term longitudinal UX measurements. Barely any of them have combined two kinds of UX in one study. To achieve sustainable development of mobile apps, we agree with researchers who emphasize positive long-term user experience. However, we propose that a great first impression matters as well. This is especially the case for mobile apps. Research shows that 25% of new users will log in to new apps only once, and 80% of new users will decide whether to continue using the app within three minutes of use [18]. Similarly, Flurry, an iPhone app metrics company, found that a free iPhone application loses 95% of users after one month [18].

Nevertheless, very few studies have evaluated the first-time user experience, despite its importance to the success of the product [19]. To this end, we developed this study to assess users’ first impressions, as well as how their relationship with the product had changed in the long-term. From our perspective, both the first-time and the long-term user experience should be considered if companies aim to select the right design options from UX evaluation feedback. A great first impression helps to attract users, and a positive long-term UX further assists companies to retain users. The product we chose to study was a free mobile fitness application. Free apps tend to have a shorter life cycle, which means the development teams may be under more pressure to not only attract new users constantly, but also to retain existing users.

This study is based on a multi-method approach. We began by adopting the AttrakDiff Questionnaire to assess users’ first impression of the application, and then used the UX Curve method to evaluate how users’ experience with the app changed over time. As the two methods are able to investigate various experiential dimensions, we decided to follow the AttrakDiff to focus on four dimensions: “pragmatic qualities,” “hedonic qualities—identity,” “hedonic qualities—stimulation,” and “attractiveness.” Meanwhile, we interviewed the participants after each evaluation session,

and asked them to explain which factors deteriorated their experience. Basing on collected qualitative data, we managed to identify critical issues with two separate types of user experience.

The study has the following organization: Section 2 reports on the previous works on UX measures, and is then followed by Section 3, which illustrates our research methodology, including data collection and data analysis. Section 4 demonstrates the experimental study we conducted, in addition to the results. Lastly, Section 5 presents the conclusion of this study and outlines our future work.

## 2. Literature Review

From the late 1990s, UX has gradually been adopted as a tool to evaluate the quality of human–product interactions [15,20]. The earlier studies on UX aimed to convince the Human-Computer Interaction (HCI) community to pay more attention to users' internal state [21]. For example, Alben and their colleagues first brought up the notion of UX in 1996; in their words, UX is “the way the product feels in hands, how well users understand how it works, how users feel about it while they're using it, how well the product serves users' purposes, and how well the product fits into the entire context in which users are using it” [9]. Similarly, Prof. Hassenzahl states that “UX is about technology that fulfills more than just instrumental needs,” and it should “contribute to our quality of life by designing for pleasure” [2]. In short, UX emphasizes both pragmatic qualities and hedonic qualities.

Previous UX studies have predominantly focused on short-term measurements, emphasizing experiential qualities related to improvements of product's first impression, such as usability and usefulness [11,17,22], while in recent years, an increasing number of researchers have started to emphasize the great importance of long-term user experience evaluation. Studies find that hedonic qualities like pleasure and beauty play an essential role in long-term use [23]. Positive long-term user experiences suggest that improved customer satisfaction and loyalty can foster business sustainability [16,24]. This helps to explain why there has been a shift of emphasis from “an experience” (one test session only) to a longer period of use [17,25].

Researchers and practitioners have presented various methods to assess user experience. Karapanos et al. divided UX measure methods into three mainstream perspectives: cross-sectional, pre-post longitudinal, and retrospective reconstruction [26]. Cross-sectional approaches distinguish users by different levels of expertise, or the length of time users use the product [26]. Such approaches neglect that external variation is not entirely controllable, but attribute variation across users to the manipulated variables [27]. This was supported by the study by Prümper et al., which found that different definitions of define “novice” and “expert” users led to different results [28]. In this case, it is hard to say how close the research result is to the real user experience.

Pre-post approaches study the same users at two points in time [26]. For instance, Karapanos et al. invited ten users to test a novel pointing device and record their user experience of the product during their first experiences with the product, as well as after four weeks of use [29]. The limitation of pre-post approaches is that only two measurements are made of the user over an extended period of time, so time effects may not be the only reason for the changes in their experience with the product [27]. Moreover, user experience changes through the process of interactions; merely assessing a short-term time frame or a few points in time is unlikely to provide precise information to forecast a fully comprehensive user experience in real life [30]. There are few longitudinal studies covering several months or even the life cycle of a product [27,31], as they are expensive, impractical, and trained researchers are often required [26]. Therefore, even companies that value the importance of long-term longitudinal UX measurement can seldom employ this approach.

Retrospective approaches support users to recall the most memorable experience they had with the product within a given time period [27,32]. These approaches derive from the CTI (critical incident technique). In CTI studies, participants are asked to reconstruct meaningful personal experiences based on memory. Von Wilamowitz-Moellendorff et al. proposed change-oriented analysis of the relationship between product and user (CORPUS), a structured interview technique to retrospectively evaluate users' perceptions of different aspects of perceived product quality [33]. Participants in the CORPUS

are asked to compare their current attitude of the product to the moment when they purchased the product, and rank a few UX dimensions on a 10 point scale over the timelines. If there is any change in their experience, participants need to explain the trends and causes of changes further.

Kahneman et al. proposed the experience sampling method (ESM), a retrospective diary method to collect users' experiences "in situ" [34]. The in situ data collection can minimize the bias caused by memory effects, but it requires high compliance from participants, as well as rather high effort for participants. Building off the ESM, researchers further developed the day reconstruction method (DRM), which imposes a chronological process in the daily retrospective assessment. Instead of requiring participants to report all use cases with the product immediately, which might interrupt their ongoing activities, the DRM asks participants to reconstruct a few of the most impactful emotional experiences of the preceding day. Nevertheless, due to their time-consuming and laborious nature, long-term longitudinal studies like the ESM and DRM are still rare [35].

There are examples in the literature of lightweight methods for evaluating long-term UX. Compared to interviews and transcripts, graphing is assumed to be a more straightforward method. Sonnemans and Frijda first adopted graphing to recall past emotional experiences in 1994 [36]. Karapanos et al. presented iScale, an online survey tool for supporting users to retrospectively reconstruct their experience with a product by sketching curves from the moment of purchase up until the present [26]. The pioneer study by Karapanos et al. showed that participants using the iScale tool were able to recall more experience reports than those who reported directly without any form of graphical representation [26]. As an online survey tool, users can easily modify curves in iScale. However, comparing to a paper-and-pencil method, iScale provides a lower degree of freedom in sketching, and users may find it is harder to annotate their sketches in iScale as well. Thus, free-hand sketching is claimed to be a "more expressive" tool [30]. Moreover, iScale is not publicly available yet, so it may not suit industrial contexts where resources are limited.

Kujala et al. developed the UX Curve method, a paper-and-pencil method which also requires participants to draw several curves describing how their experience with the product has changed over time. The UX Curve involves face-to-face interviews between researchers and participants, so that participants' reasoning and thoughts can be better-collected [30]. Researchers can further study the meaningfulness of how these recalled experiences affect user satisfaction and customer loyalty [31]; this is essential for UX improvements. As Kujala et al. state, a "method is useful until it is known what the resulting data means and what separates happy and unhappy users" [30].

Some may question the validity of retrospective studies, arguing that human memory is often incomplete and biased. Admittedly, memories are rarely accurate, but they should not be underestimated. Retrospective memories suggest a significant association between users' recommendation and loyalty and product success. Shiffman et al. argue that retrospective recall has a huge effect on individuals' later behavior, as it represents the information that individuals use to make future decisions [32,37]. Psychological studies also prove that memory introduces systematic biases into individuals' evaluations, especially at the peak and end of the experience [38]. Subjective factors such as reconstructions and emotions largely influence users' decision-making and judgment [39]. For users, their perception of the product determines whether they will continue to use, purchase, and recommend it to others in the future. The evidence presents in Garrett's work suggests that customers who have a delightful experience with the product are more likely to purchase it again [40]. As Karapanos et al. pointed out, "it may not matter how good a product is objective, it is the 'subjective', the 'experienced', which matters" [26].

### 3. Research Methods

For products like mobile apps, short-term subjective evaluations are very likely to differ from UX after a longer period of use, as users may need time to learn how to use them. Therefore, merely investigating users' feelings about the app after first or a short-term use, or just focusing on the long-term user experience can both be used as a basis for predicting future product success as well as

users' future behavior [31]. Meanwhile, the existing approaches that measure longitudinal experience tend to be expensive and laborious. Thus, we aim to introduce a cost-effective UX evaluation method by combining the pre-post approach and the retrospective approach. The details are given in the next chapter.

### 3.1. The Limitations of Existing Methods

This study aims to introduce an easy to apply approach for measuring user experience of mobile apps. The importance of long-term user experience of a product has already been mentioned in the previous Sections, and we recommend the UX Curve because it has proven to be an efficient and easy-to-use UX measurement method in the long-term [30,35]. However, previous UX Curve studies have predominantly focused on products such as mobile phones [30] or Facebook [35]. These two kinds of products are necessities for many individuals, and are frequently and consistently used over time. Therefore, users are unlikely to stop using these products due to a few negative experiences. Thus, developers of these products pay more attention to long-term user experience, because a good long-term user experience is believed to improve customer loyalty [30]. For users, the cost of abandoning a phone or quitting Facebook could be high. However, this is not the case for most mobile apps. In the era of Mobile Internet, electronic devices such as mobile phones and tablets have become more and more critical in our daily lives, which has led to an upsurge in the number of mobile apps. Numerous apps can provide similar functions, not to mention the fact that a large number of them are free to download. If users feel the app is not attractive at first glance, there is a high possibility that they will stop using it and find alternatives.

We argue that users' first impression of an app plays an essential role in the success of the app, especially for free apps. While a user might not directly pay for an app, apps have costs to the companies that create them. For those free apps, daily active users (DAU) and monthly active users (MAU) are two critical predictors of monetary conversions in later use [41]. If apps lose users immediately after their first-time use, active use will be inevitably low. Additionally, the first impression has a long-lasting effect on retention, which is critical to the success of apps. Without long-term retention, app companies could lose their chances to engage and monetize users in the future. Based on these facts, we suggest that the first-time user experience holds dominant power at the early stage.

However, long-term UX is still important for the sustainability of an app. Therefore, we aimed to introduce a UX evaluation approach that can not only assess the long-term user experience, but also the initial user experience after users' first interaction with the app. Our approach also defined critical issues with two types of user experience separately. By demonstrating that users encountered different kinds of UX issues in different usage periods, we hope to reinforce that developers and designers should pay attention to two types of UX simultaneously.

The UX evaluation methods we chose to use were the AttrakDiff questionnaire and the UX Curve. The two approaches have been adopted and tested by researchers and industry people. In our study, the AttrakDiff questionnaire was used to assess users' initial experience with the app, and the UX Curve method was adopted to evaluate how users' relationships with the app changed over time.

We chose to combine these methods because they complement each other well. The UX Curve has proven to be an effective method for measuring the hedonic qualities of user experience, and it is confirmed that experiential factors such as enjoyment and pleasure play an essential role not only in long-term user experience, but also in customer loyalty [30]. However, as a retrospective user experience evaluation method, it is not suitable to assess how participants feel about the app after executing a task. Thus, the importance of the first experience to app development is neglected. Therefore, we adopted the AttrakDiff questionnaire for the first-time UX measurement session. There were mainly two reasons why chose the AttrakDiff questionnaire: 1. it was a well-established UX evaluation method, and can be applied to a wide range of products; 2. in the context of users' initial encounters with the product, the pragmatic qualities are vital for users' goodness judgments of the product [35], while questionnaires are suggested to be more suitable for identifying the pragmatic aspects of user



experience [35]. In addition, questionnaires have been suggested to be more suitable for identifying the pragmatic aspects of user experience [35].

### 3.2. Introduction of Two Methods Formats

#### 3.2.1. The AttrakDiff Questionnaire Format

In this part, we would like to introduce the AttrakDiff questionnaire. The AttrakDiff is a method developed by Hassenzahl et al. to assess the perceived pragmatic quality, the hedonic quality, and the attractiveness of the interactive product. Pragmatic quality refers to the product's usability, that is, how the user can achieve his/her goals by using the product. Hedonic quality refers to the psychological needs of the user. The AttrakDiff further divides hedonic quality into two subdimensions: stimulation and identification, which refer to the product's potential to provide positive emotional experience and ownership separately [22]. Hassenzahl indicates that hedonic and pragmatic qualities are independent of one another, but both contribute equally to the rating of attractiveness [42].

AttrakDiff evaluates the user experience by using the technique of the semantic differential on pairs of opposite adjectives (e.g., "confusing–clear," "good–bad"). For the present study, we used a Chinese version of the questionnaire translated by the authors. Adjectives were evaluated on a seven-point scale, ranging from –3 to 3, with 0 indicating neutrality [42]. It consisted of 28 items, broken down into four dimensions:

Pragmatic quality (PQ): Technical–Human; Complicated–Simple; Impractical–Practical; Cumbersome–Straightforward; Unpredictable–Predictable; Confusing–Clearly structured; Unruly–Manageable.

Hedonic—identity quality (HQ-I): Isolating–Connective; Unprofessional–Professional; Tacky–Stylish; Cheap–Premium; Alienating–Integrating; Separates me–Bring me closer; Unpresentable–Presentable.

Hedonic—stimulation quality (HQ-S): Conventional–Inventive; Unimaginative–Creative; Cautious–Bold; Conservative–Innovative; Dull–Captivating; Undemanding–Challenging; Ordinary–Novel.

Attractiveness (ATT): Unpleasant–Pleasant; Ugly–Attractive; Disagreeable–Likeable; Rejecting–Inviting; Bad–Good; Repelling–Appealing; Discouraging–Motivating

The questionnaire has been widely used by practitioners and researchers, and has been suggested to be an easy-to-use and effective UX assessment method.

During the first evaluation session, the participants were given around ten minutes to explore the application. After their first encounter with the app, they were asked to fill in the AttrakDiff questionnaire. Participants' evaluation of each UX dimension can be reflected on their rating for this dimension.

The questionnaire results highlighted problematic UX aspects, but were unable to identify specific issues. For example, if the participant scored the PQ dimension below zero, it was suggested that the participant was not satisfied with the application's pragmatic qualities, however, it was unclear what caused his/her dissatisfaction. In order to analyze issues in more depth, we added an interview session afterward. When a participant completed their questionnaire, researchers calculated values of four UX dimensions. If the participant rated any dimension a negative score, they were required to explain why their experience was negative orally or verbally. This session was audio-recorded and was transcribed before analysis.

Participants' first-time user experience measurement session ended here.

#### 3.2.2. The UX Curve Method Format

Before the UX Curve graphing session, participants were required to use the app on a daily basis for four weeks. They could choose whenever they want to use it, but they need to ensure that each use was no less than five minutes. Four weeks later, participants attended a UX Curve graphing session where they would recall important details of the application qualities that affected their experience in the four weeks.

Kujala et al. developed UX Curve, a cost-effective method that assesses long-term UX based on user-reconstructed memories [30]. According to the developers, the UX Curve is an easy to apply tool that assists users to recall significant details of the product qualities that affect UX [30]. Participants are asked to draw curves describing how their experience changed from the moment of purchase to the end in the present from different viewpoints [30]. Users determine what is meaningful to them freely, and explain the reasons why their experience has changed over time. In the pilot study of the UX Curve, researchers managed to prove that the curve trends were related to user satisfaction, user loyalty, and users' willingness to recommend the product; moreover, the underlying reasons provided by the participants can assist designers and developers to improve the product [30].

To meet the needs of our study, we made some adjustments to the original the UX Curve method. The prior UX Curve study by Kujala et al. focused on five dimensions of UX: general UX, attractiveness, ease of use, utility, and degree of usage [30], while in our study, in order to combine well with the Attrakdiff, our UX Curve method investigated the same dimensions as the questionnaire did: pragmatic quality, hedonic quality—identity, hedonic quality—stimulation, and attractiveness. Accordingly, we developed four versions of the UX Curve's template to gather data on the four dimensions. These templates were used to assist users to draw their curves presenting how their relationship and experiences changed over time. Each template included an empty two-dimensional graph area, with the horizontal axis representing time dimension, from the user's first interaction with an app to the current moment. The vertical axis represented the intensity of the user's experience. A horizontal zero line lay in the middle of the graph, which divided the area into two parts: a positive upper part and a negative lower part. Accordingly, the vertical axis was labeled with + and – signs. On the top of the template, the instruction was given as: "Please recall the moment when you first used the app, and then draw a curve to describe how your feeling of the app has changed from the beginning to the present. Also, please mark the reasons on the curve." As all the participants were Chinese, we used the Chinese language during the study.

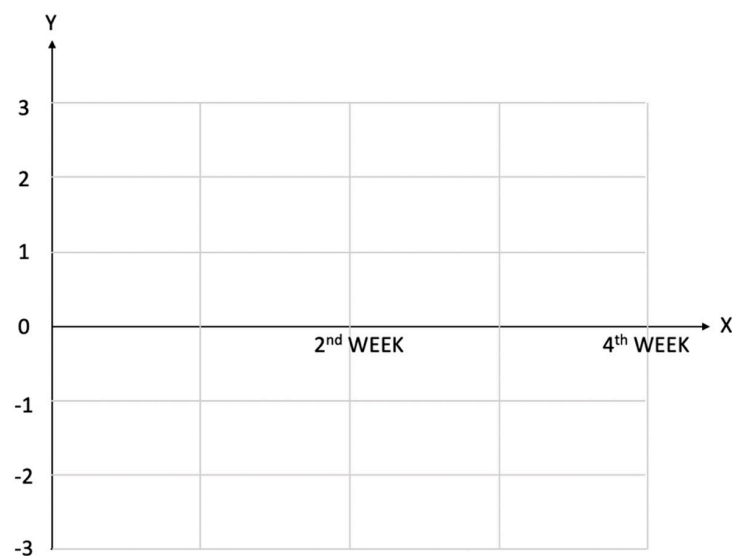
We made some modifications to the UX Curve template as well. In this study, the Y-axis consisted of a seven point scale with values from –3–3, and the values were represented in the form of emotions (–3 = very negative, –2 = moderately negative, –1 = slightly negative, 0 = neutral, 1 = slightly positive, 2 = moderately positive, 3 = very positive) (Figure 1). That is to say, participants could rank four UX quality dimensions of the app on a scale from –3–3. This helped to quantify the UX better. The original UX Curve method merely focused on the curves' trends (improving, deteriorating, or stable). We attempted to take a step forward by assessing to what extent the user's long-term experience changed, by calculating the difference between the vertical values of the ending and the starting points of the curve. If the difference was positive, the long-term experience of the user was deemed to be improved after their first use. Vice versa, if the difference was negative, the long-term experience of the user was considered to be decreased. If the difference was zero, the user was deemed to have a relatively stable experience.

As the user's initial experience with the app has been evaluated by the questionnaire in advance, we marked the scores of his/her initial experience of the app on each participant's UX Curve template, and these points were regarded as the starting points for the curves. The advantage of this move was that it showed clearly how the long-term experience of the user changed after the initial use.

Traditional UX Curve studies require participants to explain the reasons for the changes in their user experience, including positive and negative reasons. In our study, in order to focus better on domain issues as well as to reduce participants' workload, we only required participants to explain adverse changes. This interview session was audio recorded and transcribed, so participants could choose to explain their reasons verbally or write them down in the curve templates.

We then conducted a content analysis of the responses and identified the critical problems with the long-term user experience. Considering the similarity of UX issues between the HQ-I and HQ-S, we conducted a joint analysis with the data collected in the HQ-I and HQ-S aspects to reduce the analysis redundancy. That is to say, although we divided the hedonic qualities into two parts in the

earlier sessions, when we finally summarized the issues of the hedonic dimension, we combined the two subdimensions (stimulation and identification) together.

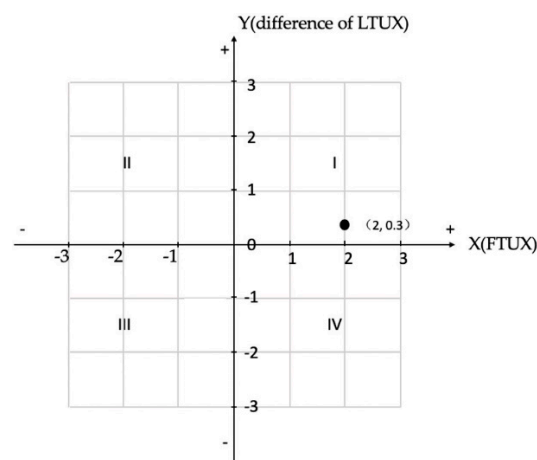


**Figure 1.** An example of our user experience (UX) Curve template.

### 3.3. Data Analysis

#### 3.3.1. Four UX Dimension Coordinate Planes

The first purpose of this study was to determine the quality of two kinds of UX, and how users' experiences with the app changed from their first interaction to after four weeks of repeated use. We developed four coordinate planes, each addressing the viewpoints specific to the UX: PQ; HQ-I, HQ-S, and ATT. In each coordinate plane, the X-axis represented the value of the specific AttrakDiff dimension, and the Y-axis represented the difference of the ending value and the starting value of the same UX Curve dimension. In other words, how the user rated his or her first-time user experience was X, and how much his/her long-term experience had changed was Y. In this case, we could graph users' experience on the coordinate planes accordingly. For example, if a user rated the HQ-I dimension a 2, and the difference between his ending HQ-I UX Curve value and the starting value was 0.3, he can be represented by (2, 0.3) on the coordinate plane (Figure 2).



**Figure 2.** An example of UX coordinate plane.



With the help of the coordinate planes, we could determine the quality of first-time UX and long-term UX. Here, we take the PQ dimension as an example; the analysis processes of the other three dimensions are similar. The distribution of points showed how the participants assessed the app. After entering participants' two types of UX values in the PQ coordinate plane, the distribution of coordinates can assist researchers to gain a rough idea about how users evaluated the FTUX and the long-term UX. The intersecting X- and Y-axis divided the coordinate plane into four quadrants (Figure 2), and we named them Quadrant I, II, III, and IV. Quadrant I, where both x and y were positive, represented a positive FTUX and an improved long-term UX. Similarly, Quadrant II represented a negative FTUX but an improved long-term UX; Quadrant III represented a negative FTUX and a decreased long-term UX, while Quadrant IV represented a negative FTUX but an improved long-term UX. When the most points were located in Quadrant I, it was suggested that most users' first impressions of the app's pragmatic qualities were positive, and they were even more satisfied with such qualities after long-term use. This was a desirable situation for app developers, and if they aimed to provide users with a better user experience, they could work on how to improve the pragmatic dimension of the first user experience. If more points lay in the second quadrant, developers should realize that although users' long-term user experience was on the rise, their initial user experience with the app tended to be poor; in this case, app developers might need to improve the user's first experience by minimizing the barriers to entry, thus retaining as many users as they can at the beginning. App developers should be alert to the situation where the most points are located in the third quadrant. This indicates that not only are the users unsatisfied with the app's practicality and functionality after their first use, they also have negative memories of the use of the app's pragmatic qualities after repeated use. If this was the case, developers would be better to figure out how to improve FTUX and long-term UX simultaneously. However, we recommend them to solve the long-term experience problems first, because it seemed that the long-term user experience was even worse than first-time user experience. Finally, if the most points lay in the Quadrant IV, app companies were suggested to invest more resources in improving the pragmatic qualities of the long-term user experience. This is because most user-app relationships saw a decreasing trend, while the users' first encounters with the app were positive in the most cases.

If developers were unable to determine which quadrant had the most points, or if they looked for a more accurate result, they could calculate the mean values of the two kinds of experience scores separately, and then define which kind of experience problems the app needs to tackle immediately, according to the quadrant in which the mean point value lies.

Notably, the ATT coordinate plane was able to determine how users evaluated the two kinds of user experience in general. According to Hassenzahl, the developer of the AttrakDiff questionnaire, hedonic and pragmatic qualities both contribute equally to the rating of attractiveness. We thus believe that designers or developers could understand what kind of FTUX and long-term UX their products can provide in general by analyzing the ATT coordinate plane.

### 3.3.2. Identifying Critical UX Issues Based on Quantitative Data

Our second aim was to identify the most critical issues of the first-time and the long-term UX separately, so that the UX measures could be used in the next cycle of app development. We defined the problems based on the information obtained from two interviews. During interviews, participants were asked to illustrate the causes of their negative experience. We conducted a content analysis and classified all the given negative reasons for similar meaning as an issue, and then counted the frequency of mention of each issue. For example, if one participant stated that "it took over three minutes to sign up" during the first interview, and another participant mentioned "the account creation process was too long," we then assigned the "FTUX-sign up" issue a "2"; the "2" stands for the number of times the issue was mentioned.

Considering that there might be many problems raised, and the severity of the problems varied, we recommend that companies target resources to address issues that most impact both types of UX.

In this study, we focused on the six issues (three for the FTUX and three for the LTUX) that were mentioned the most.

The analysis of the UX problems ended here. In the end, we identified the most serious problems of pragmatic qualities and hedonic qualities that users faced when they used the app. The developers should solve these problems first. Inevitably, all products have bugs, and all products can be improved upon; this is the process of designing and improving products. We recommend that app companies prioritize their problems according to severity, and start improvement processes from the most serious ones. This can be achieved by comparing how users evaluate the four dimensions of the two kinds of user experience.

#### 4. Experimental Section

In this chapter, we assessed the usefulness of our proposed UX assessment approach in a study with 20 participants. Our work aimed to investigate if our proposed UX assessment approach was practical, and, furthermore, whether the method could be used to identify factors deteriorating user experience of the app. The research work thus included three steps. First of all, we recruited a group of participants to evaluate their first-time user experience of the tested app. After four weeks of repeat use, these participants joined the UX Curve drawing session, where their long-term experience with the app was assessed. The last step was to analyze the data and information collected from the two UX evaluation sessions, and to conclude which were the most serious UX problems of the app.

The exploratory study was conducted in Beijing, China. We worked with a free fitness app running on Android and iOS platforms. The app was developed by a China Mobile subsidiary that specializes in mobile internet operations. As a fitness app, it targets users between the ages of 18 and 45 years, and aims to prompt users to follow through on their fitness goals by providing tools like activity tracking, video-based workout sessions, and nutritional information. The app was first launched in 2016, and its developers release an update every 6 weeks. The app remained unmodified during the 4 week study to ensure that the participants used and gave feedback on the same app.

Before the study, the UX of the fitness app was considered to be relatively poor, which can be seen from two facts: (1) Its retention rate was lower than average retention rate for mobile apps. By day 1, the app had an 18.9% retention rate, which reduced to 5% in 10 days, and then dropped to 0.87% by day 90, compared to 21.0%, 7.5%, and 1.8% for average apps [43]. (2) Ratings of the app in major app stores have been dropping for two quarters with over 19% of user reviews negative.

##### 4.1. Data Collection

###### 4.1.1. First-Time User Experience

Twenty participants were recruited for the study. Participants were selected with different backgrounds, ages, genders, and mobile phone models. We also set a few qualifications for candidates. First of all, since the study needed to measure the users' first-time user experience with the application, participants could not have used the application before. Secondly, participants were physically qualified to use the fitness application for at least five minutes on a daily basis during the study. They could explore any features the app provided at any time during the day.

The proportion of female (50%) and male (50%) participants was equal. Their ages ranged from 18 to 42 years (mean 27.2 years) and they belonged to the target group of fitness apps. All of the participants used a smartphone, though mobile phone models varied, representing five brands: iPhone, Huawei, Xiaomi, Vivo, and Oppo. Participants used their smartphones to access the Internet every day. The personal features of participants are shown in Table 1.

**Table 1.** Participants' basic features (user content needs survey).

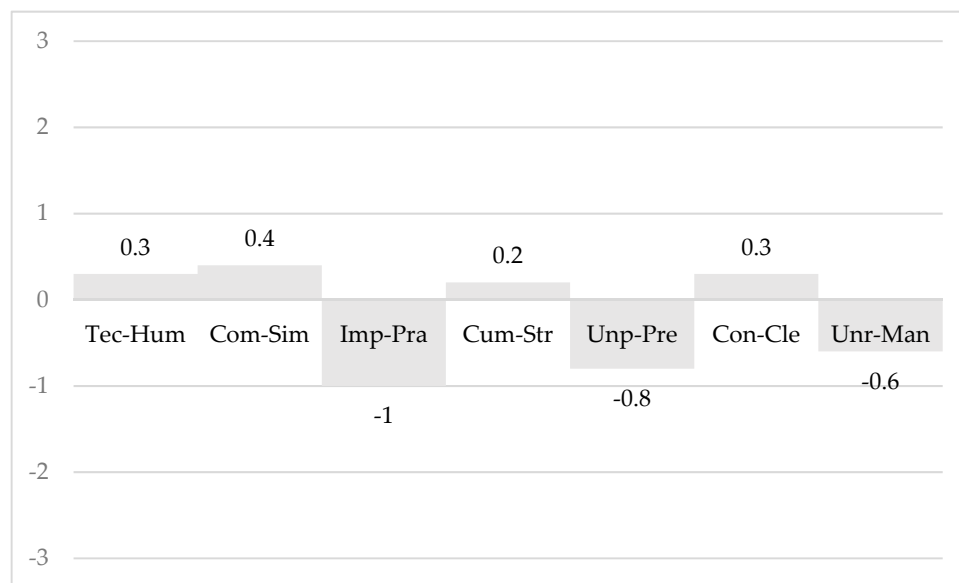
Feature	Participant Distribution
Gender	50.0% Female 50.0% Male
Age	10.0% Under 20 years old 25.0% 20–24 years old 30.0% 25–29 years old 25.0% 30–34 years old 5.0% 35–39 years old 5.0% 40–44 years old
Education	15.0% High school 50.0% Bachelor's degree 30.0% Master's degree 5.0% Other (Ph. D, Professional Course, Uneducated, etc.)
Occupation	60.0% Employee 10.0% Freelance 20.0% Student 10.0% Unemployed
Phone OS	30.0% iOS 70.0% Android

The evaluation sessions with the participants consisted of an AttraDiff questionnaire and a curve-drawing session, as well as two interviews.

The 20 participants joined the AttraDiff questionnaire session on August 1, 2018. After using the app for ten minutes, participants filled out the questionnaire, which aimed to assess their first-time experiences with the app. The collected data are reported in Table 2. The HQ-S dimension was the only dimension in which the mean value was greater than zero (0.6). The other three dimensions were rated between  $-1$ – $0$ . HQ-I was scored with the second highest mean value,  $-0.3$ , followed by the ATT aspect ( $-0.4$ ). It seems that participants did not enjoy their first interaction with the application much, especially its pragmatic qualities, which only scored  $-0.6$ ; this indicated that the PQ dimension of the first-time user experience was most problematic. Thus, we decided to take a closer look at the evaluation of specific items contained in the PQ dimension. Figure 3 provides the average values of seven items contained in the PQ aspect. The assessment of the item Impractical–Practical obtained the lowest score ( $-1$ ). The other two pairs of items that were rated below zero were “Unpredictable–Predictable” and “Unruly–Manageable,” which scored  $-0.8$  and  $-0.6$  respectively. It is suggested that the app was unsuccessful in satisfying participants' practical and functional needs when they first tried the application.

**Table 2.** The AttraDiff questionnaire results (user perception survey).

Score Band	PQ	HQ-I	HQ-S	ATT
−3	2	1	0	1
−2	4	4	2	5
−1	5	5	3	5
0	4	4	4	3
1	3	3	5	3
2	2	2	5	2
3	0	1	1	1
Mean Value	−0.6 (s = 1.5)	−0.3 (s = 1.6)	0.6 (s = 1.4)	−0.4 (2 = 1.6)



**Figure 3.** The average values of seven items contained in the Pragmatic Quality (PQ) aspect.

After filling out the AttraDiff questionnaire, participants had a break for 30 minutes before the first interview session, where they were required to explain why they were unsatisfied with the app after their first use. The reasons were content-analyzed and further categorized into pragmatic and hedonic issues. We used the user experience model by Hassenzahl for categorization [6]. We collected 25 first-time user experience issues in total from participants, 15 for pragmatic issues, and 10 for hedonic issues. Table 3 represents the categories and three most-mentioned issues related to each category. The pragmatic issues were mostly related to technical faults and bugs (mentioned five times). According to collected data, crashes were the most serious issue that impacted user experience. One participant mentioned that the application had crashed twice in eight minutes, and that he would uninstall the app straight away if he was not involved in this study. Another critical issue was associated with “registration process”; four participants complained about it. Three participants mentioned that it would be better if the app could connect with their fitness trackers.

**Table 3.** Issue list of the first-time UX.

The First-Time UX	
Pragmatic Issues	Hedonic Issues
Crash bugs (5 times)	Social sharing (4 times)
Long registration process (4 times)	Limited workout options (3 times)
Restricted Connection (3 times)	Unpleasant designs (2 times)

The hedonic reasons (including the HQ-I and the HQ-S issues) largely referred to identification and beauty; four participants complained that, compared to other popular fitness apps, the test app was more “isolated,” since they could not share their performance and achievement with friends. Three participants scored the hedonic qualities dimension below zero because there were not enough workout options, so they described the application as “unprofessional.” Two participants stated that they did not like the design and visual appearance of the application.

#### 4.1.2. The Long-Term User Experience Evaluation

Four weeks later, the twenty participants joined the second evaluation session, in which they reconstructed how their relationship and user experience with the app had changed during four weeks of use by drawing curves. We collected four different curves with the four curve templates from each user, so a total of 80 curves were collected, including 20 for each of the four UX dimensions. The curve forms were categorized according to the difference between the vertical values of the ending and starting points of the curve to positive, negative, or zero. The results are reported in Table 4. In total, most of the differences were positive, implying an improving long-term user experience, but the PQ dimension had more negative differences than positive ones, suggesting a deteriorated experience of the PQ aspect. More than a half of the HQ-I differences were positive, which means that most participants became more satisfied with the hedonic quality—identity. Half of the participants also had a better experience of the hedonic quality—stimulation. This dimension had the least negative difference as well; only 25% of participants drew decreasing HQ-S curves. As for the ATT dimension, 65% of participants graphed an improving or stable curve, suggesting that most participants had an improving or stable user experience in the four weeks.

**Table 4.** The number of the difference of the different UX dimensions.

Difference	PQ	HQ-I	HQ-S	ATT
Positive	7 (35%)	11 (55%)	10 (50%)	9 (45%)
Negative	10(50%)	6 (30%)	5 (25%)	7 (35%)
Zero	3 (65%)	3 (15%)	5 (25%)	4 (20%)

After the curve drawing session, we had the second interviews with participants regarding what caused negative experiences during their four weeks of use. Some reasons have been marked on their curve templates. We picked out all the negative reasons that participants stated during the second interview session, as well as those written on the UX Curve templates. In total, we collected 39 reasons and then categorized them according to Hassenzahl’s model into pragmatic and hedonic issues. Of collected reasons, 25 related to pragmatic qualities and 14 were hedonic issues. In detail, the most mentioned long-term UX issues were bug crashes (eight times), followed by “unable to sync activity data” (five times). Four participants complained about the “counting function” while explaining why there was a decrease in their long-term user experience. Other issues related to the pragmatic aspect were “the app drains the battery” (mentioned by three), “unable to incorporate music into the app” (three times), and “too many ads” (two times).

As for hedonic issues, five participants stated that “there were too many notifications”. One participant described the notifications for being “very annoying”; “I almost want to uninstall the app when I received its notifications three times in one day.” Another five participants hoped the app could provide a larger range of workouts. “I even considered paying for a Premium subscription for more workout guides,” said one of the participants. Four negative reasons were associated with “social sharing,” and one participant worried about “information security.” Details show in Table 5.

**Table 5.** Issue list of the long-term UX.

The Long-Term UX	
Pragmatic Issues	Hedonic Issues
Crash bugs (8 times)	Too much notification (5 times)
Unable to sync activity data (5 times)	Limited workout options (5 times)
Battery drain (4 times)	Restricted social sharing (4 times)

We compared the three pragmatic issues and the three hedonic issues that were most frequently mentioned in the two interviews and found that more pragmatic issues were reported than hedonic



issues. Apparently, “crash bugs” was the most critical issue of the two types of user experience. Another technical bug that bothered participants in the long-term was “unable to sync activity data”. The issue “limited workout choices” was mentioned in two interviews as well. For users, they expected to experience more functions that enabled them to go through physical transformations. Participants also valued social features, as the “restricted social sharing” issue was also mentioned during two interviews. Indeed, with the development of the mobile internet, many individuals have the need to share their activities on social networks. Though participants showed an interest in connecting to and sharing with others, they did not want to be sent too many notifications.

There were a few issues that we believe can be addressed easily, and which could help improve user experience quickly, for example, providing a simpler and faster registration process. There were two problems that were difficult to fix, namely connecting to other sports trackers and battery drain. Interestingly, when asked to explain why they were not satisfied with the app, participants tended to compare this application with a few popular fitness apps that they had used before, and told us the difference between two apps. For example, one participant mentioned that “compared to Nike Training, this app has limited workout options, while I can tailor my workout to suit my ability, fitness level, and strengths with Nike Training.”

#### 4.2. UX Coordinate Planes

We entered two kinds of user experience data for each user into UX coordinate planes. As mentioned before, the X-axis of the coordinate represented the value of the specific AttrakDiff dimension, and the Y-axis represented the difference of vertical values of the ending and starting point of the curve. Figures 4–7 show Pragmatic Quality, Hedonic Quality—Identity, Hedonic Quality—Stimulation, and Attractiveness respectively.

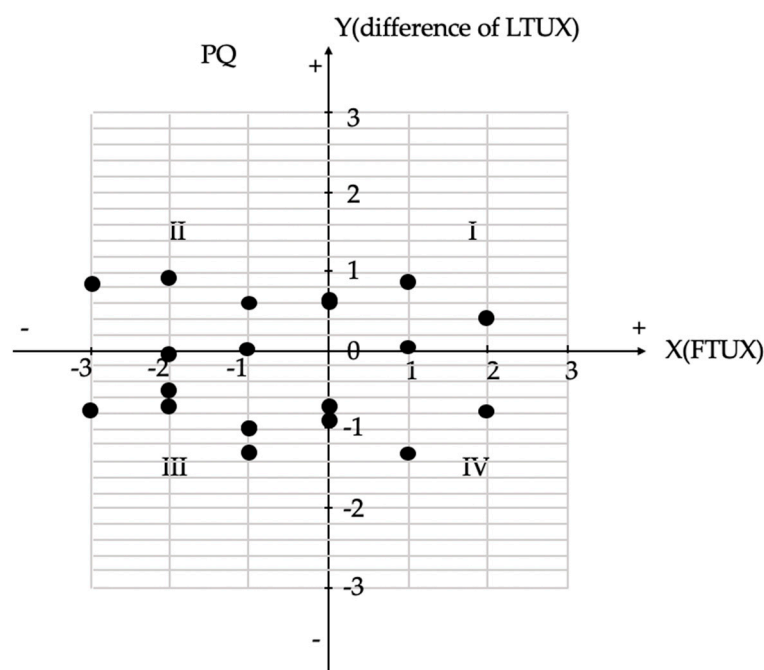


Figure 4. Pragmatic Quality (PQ) Coordinate plane results.

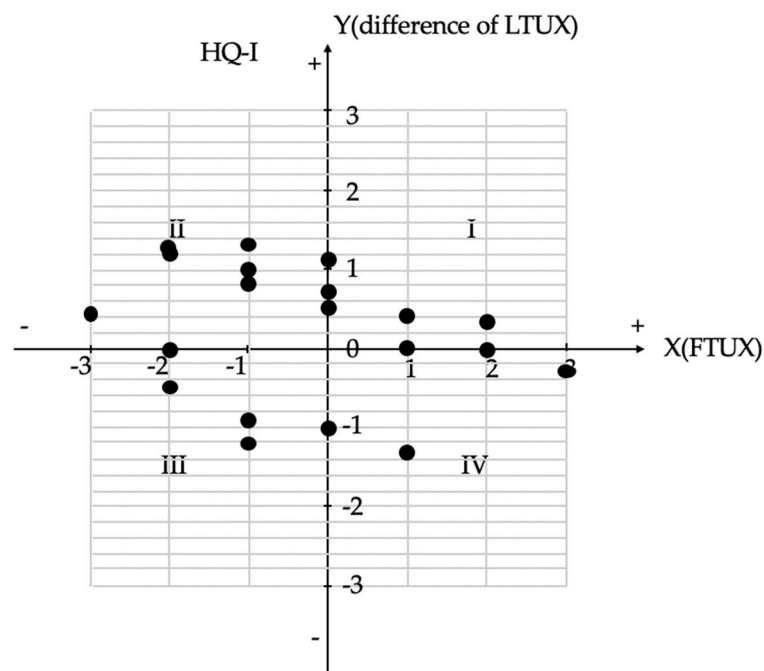


Figure 5. Hedonic Quality—Identity (HQ-I) Coordinate plane results.

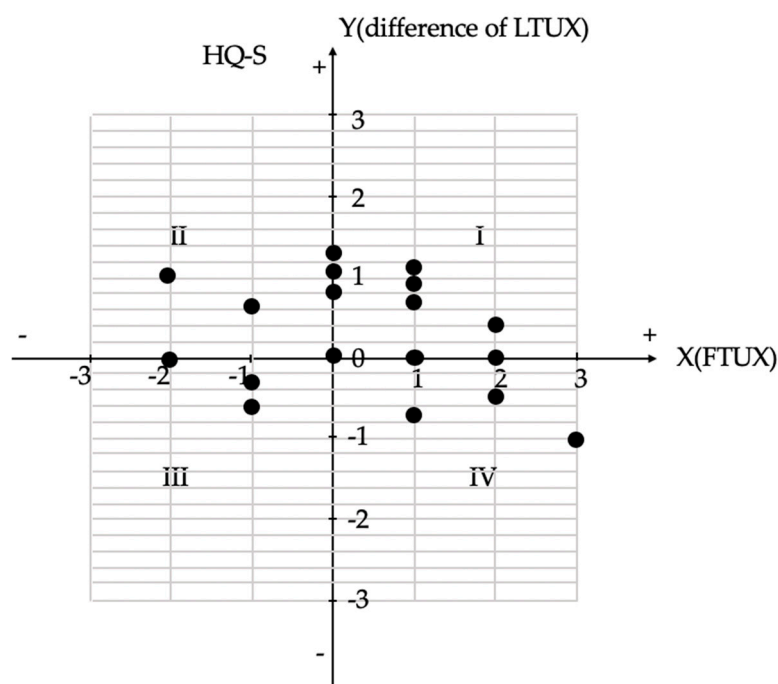


Figure 6. Hedonic Quality—Stimulation (HQ-S) Coordinate plane results.

The PQ coordinate plane (Figure 4) showed that most points were located in Quadrant III, which means that most participants were not satisfied with the app's usability and usefulness after the first encounter, and their long-term UX became even poorer. This was consistent with the pragmatic issues they reported during two interviews. Crash bugs were suggested to be the most negative issue that deteriorated both the first-time and long-term user experience.

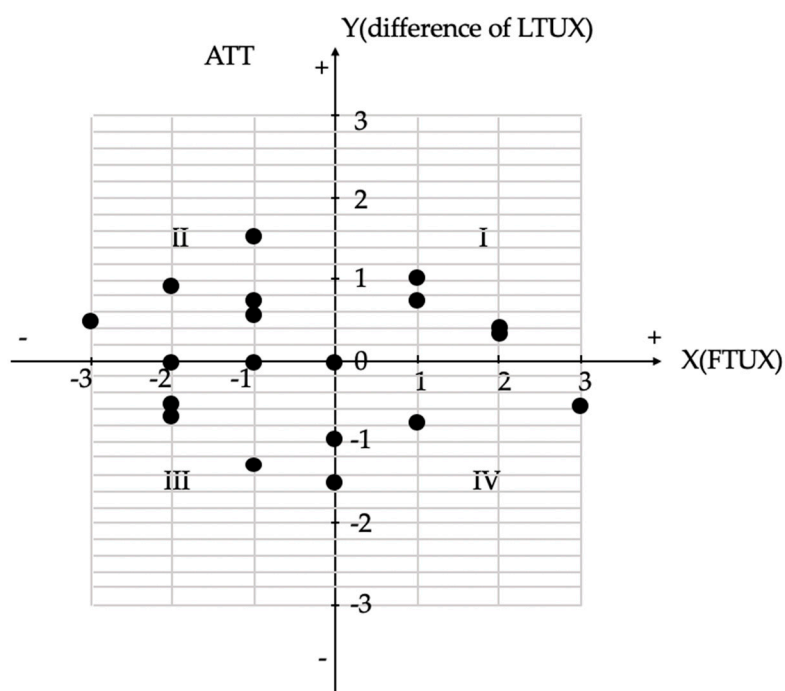


Figure 7. Attractiveness (ATT) Coordinate plane results.

As for the app's hedonic qualities, according to the HQ-I coordinate plane, most points were located in Quadrant II, indicating that participants had a negative experience with the application during the first encounter, but their experience improved after long-term use. As for the other hedonic qualities dimension, most points in the HQ-S coordinate plane were located in Quadrant I, suggesting that participants held a positive attitude towards the app's hedonic—stimulation qualities, and their user experience of this dimension even saw an increase. In generally, participants' evaluations of the hedonic aspects were better than their evaluations for the pragmatic qualities. Especially in HQ-S, for which the dimension not only obtained the highest values during the first-time UX evaluation, but also provided an improved experience in the long-term. It was suggested that the app was successful in stimulating participants. Regarding the hedonic quality—identity, limited workout options (five mentions) and social isolation (four mentions) resulted in a negative first impression, and these issues also had long-term effects on user experience.

Finally, as shown in Figure 7, most points in the ATT coordinate plane were located in Quadrant II, which means that participants might initially think that the application was not attractive, but the perceived attractiveness increased over time.

Overall, we collected more UX issues when evaluating the long-term user experience, but, interestingly, and there was an increasing trend of the user experience for three dimensions. It is necessary for the development team to pay great attention to the pragmatic issues. We also collected problems from the FTUX evaluation. Some of these problems were unable to be detected from a single long-term experience evaluation, but were likely to cause users to abandon their continued use: long registration process, for example.

## 5. Discussion and Conclusions

In this study, we aimed to adopt two different methods to evaluate the first-time and long-term user experience separately. User experience studies have previously largely focused on short-term evaluation. In recent years, a few UX studies have claimed the importance of positive long-term user experience, especially its contribution to user satisfaction, loyalty, and the company commercial success [30]. We agree on that; however, we also suggest developers pay attention to users' initial experiences with the product, particularly for mobile apps. Apps are suggested to have shorter life

cycle; evidence shows that 25% of users will log into new apps only once, and 80% of first-time users will decide whether to continue using the application within three minutes of use [18]. Unfortunately, we have not seen any UX research that has emphasized the importance of first-time user experience, nor have we seen research that has studied both the first-time and the long-term experience. To this end, we developed this study, which involved 20 participants to evaluate the two types of user experience of a fitness app. The first-time experience was evaluated using the AttraDiff questionnaire, while after four weeks' use, participants were asked to describe their experiences by means of the UX Curve method. We obtained both quantitative and qualitative data from two evaluation sessions. The quantitative data assisted us in determining qualities of two types of UX. In order to analyze how the user experience had evolved in four weeks, we developed four UX coordinate planes for four different UX dimensions, namely, Pragmatic Quality, Hedonic Quality—Identity, Hedonic Quality—Stimulation, and Attractiveness. In each coordinate plane, the X-axis represented the value of the specific AttraDiff dimension, and the Y-axis represented the difference of vertical values of the ending and starting point of the curve.

Qualitative data were collected from interviews with participants. Each evaluation session included interviews with participants, where participants could provide more feedback that enabled developers to fix experiential problems. Participants were required to explain why they had negative experiences when using the app, and we categorized their comments according to Hassenzahl's model into pragmatic and hedonic issues [6]. For each UX evaluation, we picked out three most-mentioned issues of pragmatic and hedonic aspects, so, in total, we concluded with 12 critical issues that deteriorated the user experience. It was suggested that application developers and designers solve these problems first. Therefore, we recommend that app companies prioritize their problems according to the severity of the problems. This can be achieved by comparing how users rated the four dimensions of the two kinds of user experience. Being able to set a priority of user experience improvement is essential for application development teams with limited resources.

This study might provide new insights into user experience evaluation, but also raises challenges for further research. We realize the potential for our continuous improvement. For example, we only focused on the issues that damaged user experience in this study, while during the process of measurement, we found that the user experiences of three dimensions were improved after four weeks. If developers can understand what factors can improve the user experience, they could further optimize user experience of the application. Therefore, we will continue to investigate the factors that will improve user experience in our future work. We also hope to work with application development teams to apply the usefulness of the evaluation results for user experience design.

**Author Contributions:** Formal analysis, L.F.; Methodology, L.F.; Supervision, W.W.; Writing—original draft, L.F.; Writing—review & editing, W.W.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Pine, B.J.; Gilmore, J.H. Gilmore. Welcome to the Experience Economy. *Harvard Bus. Rev.* **1998**, *76*, 97–105. [[CrossRef](#)]
2. Hassenzahl, M.; Tractinsky, N. User Experience—A Research Agenda. *Behav. Inf. Technol.* **2006**, *25*, 91–97. [[CrossRef](#)]
3. Lee, S.; Choi, J. Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. *Int. J. Hum.-Comput. Stud.* **2017**, *103*, 95–105. [[CrossRef](#)]
4. Feng, L.; Sun, B.; Wang, K.; Tsai, S.B. An Empirical Study on the Design of Digital Content Products from a Big Data Perspective. *Sustainability* **2018**, *10*, 3092. [[CrossRef](#)]
5. Manresa-Yee, C.; Ponsa, P.; Varona, J.; Perales, F.J. User experience to improve the usability of a vision-based interface. *Interact. Comput.* **2010**, *22*, 594–605. [[CrossRef](#)]

6. Hassenzahl, M. The Thing and I: Understanding the Relationship Between User and Product. In *Funology: From Usability to Enjoyment*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2003; pp. 31–42.
7. Law, E.L.C.; Van Schaik, P.; Roto, V. Attitudes towards user experience (UX) measurement. *Int. J. Hum.-Comput. Stud.* **2014**, *72*, 526–541. [[CrossRef](#)]
8. Bruseberg, A.; McDonagh-Philp, D. New product development by eliciting user experience and aspirations. *Int. J. Hum.-Comput. Stud.* **2001**, *55*, 435–452. [[CrossRef](#)]
9. Alben, L. Quality of Experience: Defining the Criteria for Effective Interaction Design. *Interactions* **1996**, *3*, 11–15. [[CrossRef](#)]
10. Van Schaik, P.; Ling, J. Modelling user experience with web sites: Usability, hedonic value, beauty and goodness. *Interact. Comput.* **2008**, *20*, 419–432. [[CrossRef](#)]
11. Bosley, J.J. Creating a Short Usability Metric for User Experience (UMUX) Scale. *Interact. Comput.* **2013**, *25*, 317–319. [[CrossRef](#)]
12. Law, E.L.C.; Roto, V.; Hassenzahl, M.; Vermeeren, A.P.; Kort, J. Understanding, scoping and defining user experience. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, MA, USA, 4–9 April 2009; pp. 719–728. [[CrossRef](#)]
13. Spohrer, J.; Stein, M. User Experience in the Pervasive Computing Age. *IEEE Multimed.* **2000**, *7*, 12–17. [[CrossRef](#)]
14. Wilson, G.M.; Angela Sasse, M. From doing to being: Getting closer to the user experience. *Interact. Comput.* **2004**, *16*, 697–705. [[CrossRef](#)]
15. Norman, D.A.; Nielsen, J. *The Definition of User Experience (UX)*; Nielsen Norman Group Publication: Fremont, CA, USA, 2016. Available online: <https://www.nngroup.com/articles/definition-user-experience/> (accessed on 12 April 2019).
16. Forlizzi, J.; Battarbee, K. Understanding experience in interactive systems. In Proceedings of the 5th conference on Designing interactive systems: Processes, practices, methods, and techniques, Cambridge, MA, USA, 1–4 August 2004; pp. 261–268. [[CrossRef](#)]
17. Bargas-Avila, J.A.; Hornbæk, K. Old wine in new bottles or novel challenges: A critical analysis of empirical studies of user experience. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada, 7–12 May 2011. [[CrossRef](#)]
18. Gençer, M.; Bilgin, G.; Zan, Ö.; Voyvodaoglu, T. A new framework for increasing user engagement in mobile applications using machine learning techniques. In Proceedings of the International Conference on Design, Las Vegas, NV, USA, 21–26 July 2013. [[CrossRef](#)]
19. Park, J.; Han, S.H.; Kim, H.K.; Oh, S.; Moon, H. Modeling user experience: A case study on a mobile device. *Int. J. Ind. Ergon.* **2013**, *43*, 187–196. [[CrossRef](#)]
20. Law, E.L.C.; Abrahão, S. Interplay between User Experience (UX) evaluation and system development. *Int. J. Hum.-Comput. Stud.* **2014**, *72*, 523–525. [[CrossRef](#)]
21. Olaverri-Monreal, C.; Hasan, A.E.; Bengler, K. Semi-automatic user stories generation: To measure user experience. In Proceedings of the 8th Iberian Conference on Information Systems and Technologies (CISTI), Lisboa, Portugal, 19–22 June 2013.
22. Marti, P.; Iacono, I. Anticipated, momentary, episodic, remembered: The many facets of User eXperience. In Proceedings of the Federated Conference on Computer Science and Information Systems (FedCSIS), Gdansk, Poland, 11–14 September 2016.
23. Ardito, C.; Buono, P.; Caivano, D.; Costabile, M.F.; Lanzilotti, R. Investigating and promoting UX practice in industry: An experimental study. *Int. J. Hum.-Comput. Stud.* **2014**, *72*, 542–551. [[CrossRef](#)]
24. Subramanya, S.R.; Yi, B.K. Enhancing the User Experience in Mobile Phones. *Computer* **2007**, *40*, 114–117. [[CrossRef](#)]
25. Aizpurua, A.; Harper, S.; Vigo, M. Exploring the relationship between web accessibility and user experience. *Int. J. Hum.-Comput. Stud.* **2016**, *91*, 13–23. [[CrossRef](#)]
26. Karapanos, E.; Martens, J.B.; Hassenzahl, M. Reconstructing experiences with iScale. *Int. J. Hum.-Comput. Stud.* **2012**, *70*, 849–865. [[CrossRef](#)]
27. Karapanos, E.; Martens, J.B.O.; Hassenzahl, M. On the Retrospective Assessment of Users' Experiences Over Time: Memory or Actuality? In Proceedings of the CHI 10 Extended Abstracts on Human Factors in Computing Systems, Atlanta, GA, USA, 10–15 April 2010. [[CrossRef](#)]



28. Prümper, J.O.C.H.E.N.; Zapf, D.; Brodbeck, F.C.; Frese, M. Some surprising differences between novice and expert errors in computerized office work. *Behav. Inf. Technol.* **1992**, *11*, 10. [[CrossRef](#)]
29. Karapanos, E.; Zimmerman, J.; Forlizzi, J.; Martens, J.B. User experience over time. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, MA, USA, 4–9 April 2008. [[CrossRef](#)]
30. Kujala, S.; Roto, V.; Väänänen-Vainio-Mattila, K.; Karapanos, E.; Sinnelä, A. UX Curve: A method for evaluating long-term user experience. *Interact. Comput.* **2011**, *23*, 473–483. [[CrossRef](#)]
31. Kujala, S.; Miron-Shatz, T. The evolving role of expectations in long-term user experience. In Proceedings of the 19th International Academic Mindtrek Conference, Ampere, Finland, 22–24 September 2015. [[CrossRef](#)]
32. Hassenzahl, M.; Ullrich, D. To do or not to do: Differences in user experience and retrospective judgments depending on the presence or absence of instrumental goals. *Interact. Comput.* **2007**, *19*, 429–437. [[CrossRef](#)]
33. Von Wilamowitz-Moellendorff, M.; Hassenzahl, M.; Platz, A. Dynamics of user experience: How the perceived quality of mobile phones changes over time. In Proceedings of the Workshop at the Fourth Nordic Conference on Human-Computer Interaction, Oslo, Norway, 14 October 2006.
34. Kahneman, D.; Krueger, A.B.; Schkade, D.A.; Schwarz, N.; Stone, A.A. A survey method for characterizing daily life experience: The day reconstruction method. *Science* **2004**, *306*, 1776–1780. [[CrossRef](#)]
35. Kujala, S.; Roto, V.; Väänänen-Vainio-Mattila, K.; Sinnelä, A. Identifying hedonic factors in long-term user experience. In Proceedings of the Conference on Designing Pleasurable Products & Interfaces, Milano, Italy, 22–25 June 2011. [[CrossRef](#)]
36. Sonnemans, J.; Frijda, N.H. The structure of subjective emotional intensity. *Cogn. Emot.* **1995**, *8*, 329–350. [[CrossRef](#)]
37. Shiffman, S.; Stone, A.A.; Hufford, M.R. Ecological Momentary Assessment. *Annu. Rev. Clin. Psychol.* **2008**, *4*, 1–32. [[CrossRef](#)]
38. Hsee, C.K.; Hastie, R. Decision and experience: Why don't we choose what makes us happy? *Trends Cogn. Sci.* **2006**, *10*, 31–37. [[CrossRef](#)]
39. Mandryk, R.L.; Inkpen, K.M.; Calvert, T.W. Using psychophysiological techniques to measure user experience with entertainment technologies. *Behav. Inf. Technol.* **2006**, *25*, 141–158. [[CrossRef](#)]
40. Garrett, J.J. Customer Loyalty and the Elements of User Experience. *Des. Manag. Rev.* **2010**, *17*, 35–39. [[CrossRef](#)]
41. Montague, E.; Xu, J. Understanding Active and Passive Users: The Effects of an Active User Using Normal, Hard and Unreliable Technologies on User Assessment of Trust in Technology and Co-User. *Appl. Ergon.* **2012**, *43*, 702–712. [[CrossRef](#)]
42. Hassenzahl, M. The interplay of beauty, goodness, and usability in interactive products. *Hum. Comput. Interact.* **2004**, *19*, 319–349. [[CrossRef](#)]
43. Leanplum. *Retention Revealed*; Leanplum Publication. 2016. Available online: <https://www.leanplum.com/trends/retention-revealed/> (accessed on 12 April 2019).

