ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)
https://doi.org/10.17559/TV-20221128070431
Original scientific paper

Effects of Dynamic Icons on the Perceived Usability in the Travel App

Yan LI, Yu Liang GUO*, Wei WEI*

Abstract: With the advancements in hardware and information technology, the use of dynamic icons in travel app interfaces to convey information has become increasingly common. The goal is to provide users with a novel experience that was not available in the static interfaces of the past. However, it remains unclear whether the availability and utilization of dynamic icons actually enhance the usability of apps and if it is done with genuine consideration for users' benefit. This study aims to investigate how users perceive and use dynamic icons in the context of human-computer interaction, using a travel app as the research object. We propose a user-centered subjective and objective hybrid evaluation methodology to examine the perceived usability of dynamic icons, taking into account different evaluation characteristics at various usability levels. Our results indicate no significant statistical difference between the perceived usability of dynamic icons and static icons. We employed two indicators, perceived ease and subjective priority, to evaluate the perceived usability of dynamic icons. The findings show that the perceived ease of dynamic icons is higher than that of static icons, with an increase of 3.72% (I) and 3.09% (II) in respective studies, but with a reduction of 13.19% compared to static icons in study III. We conclude that although the perceived usability of dynamic icons may not be significantly different from static icons, they exhibit higher objective usability in users' decision-making behavior.

Keywords: dynamic icon; interface design; perceived usability; travel app; usability

1 INTRODUCTION

With the growth of the tourism industry, travel apps have gained a significant presence on the mobile web and have garnered attention from the industry. These apps offer a diverse range of information and content, but the interface design style can be inconsistent compared to other apps such as messengers or media players. As travel apps are mainly used by the younger generation, the current priority is to design the interface in a way that is easily adaptable and allows for quick access to information.

Despite the abundance of travel apps available, the practical value and continued usage by users greatly depend on the interface design [1]. The interface of a travel app is the primary point of interaction between users and the app, making the design of the interface a crucial aspect. To satisfy user needs, it is important to adopt user-centered design thinking and usability-based design in the interface configuration. By prioritizing usability, the product can be more appealing to users and increase its market reach.

Icons are graphic symbols with meanings that convey a lot of information quickly and are easy to remember. Dynamic icons refer to icons with motion graphics that have a specific feature of optimizing interface manipulation through consistent design between actions. The app subsystem is accessed through dynamic icons in the interface, and this change in navigation design has a significant impact on the system. The use of dynamic icons in the interface design of travel apps affects the perceived usability of the system. They have a specific feature of interface manipulation, which optimizes manipulation through consistent design between actions.

2 RELATED WORKS

Designers utilize dynamic icons to enhance their designs, but whether or not a mobile travel app is superior should be based on its overall system design [2]. Designers often utilize dynamic icons with eye-catching designs to draw users' attention and achieve their goals. However, excessive use of dynamic icons can have an adverse effect as they may not have an interactive function during actual use, leading to perceptual usability problems.

Perceptual usability refers to the user's assessment of the product's usability from their own perspective [3]. It is defined as perceived usability and is comprised of factors such as user's congruence in completing tasks in the system, high efficiency, high productivity, organization, ease of use, and intuitiveness [4]. Perceptual usability reflects users' subjective experiences with the product more accurately and avoids bias from usability experts, making it a crucial component of the advanced structure of usability [5, 6]. It is a subjective aspect that constitutes the standard concept of constructing objective efficiency, effectiveness, and usability [7]. Research has shown that the easier a user interface is to use, the more often learners will use the system. Conversely, if the system is deemed difficult to use, learners will spend more time learning the system than focusing on the content [8]. Perceptual usability and ease of use can influence behavioral intentions and actual behaviors [9]. The relationship between perceptual usability, perceived ease, and actual behavior is also a significant issue in the study.

This study compared and assessed the usability of travel apps in terms of both the overall system and the perceptual usability of travel apps among travel app subsystems. Due to the attention-grabbing qualities of dynamic icons, it was focused on evaluating the impact of these icons on the usability of travel apps.

3 METHODOLOGY

3.1 Experiment Preparation

The methodology of this study encompasses the following steps: sample selection, experimental design, implementation of the experiment, data collection, and data analysis. The sample selection was performed through a combination of induction and Delphi techniques, utilizing the results of user interviews.

3.1.1 Selection of Research Subjects

The present study aimed to examine the usage patterns of travel applications available in app stores. A user-interview-based approach was employed to identify the types of applications most frequently used by users, which were found to be system tools, social network services (SNS), video, and shopping. The data from iResearch indicated that the proportion of travel app users under the age of 30 increased by 14.6% from 46.7% in January 2019 to 61.3% in June 2021 [10]. This finding highlights the significance of the younger generation, particularly those in their 20s, as a major consumer group in the tourism industry and a potential user group of travel applications.

The data from ASO 100 revealed that a total of 240 travel apps can be found on iOS, 97 on Huawei, and 68 on Baidu for Android platforms. In selecting the study subjects, a stratified sampling method was utilized, in which the travel apps were stratified into upper and lower classes based on their download ranking in the app store and fulfilling the four criteria of providing professional travel services, targeting individual travel consumers, being accessible and available. The characteristics of the selected travel apps were analyzed using various indicators, including founder background, release time, update frequency, latest version, size, total number of downloads, app rating, average rating, and number of reviews.

In this study, the Delphi method was employed to analyze the selected sample in a comprehensive and systematic manner. The method involves multiple rounds of informational feedback from experts and research participants, with the aim of achieving an objective and reliable result that reflects the consensus of the expert group. The open-ended format of the first expert review provided a wide range of perspectives on the evaluation criteria of travel apps, including aspects such as design, service, and marketing strategies. These insights were synthesized with the principles of usability to construct a scale questionnaire for the secondary research study. The questionnaire consisted of 9 items rated on a 5-point Likert scale, and the results of the first survey are shown in Tab. 1. This study aimed to explore the evaluation criteria for travel apps in a professional and in-depth manner, leveraging the Delphi method to achieve a comprehensive and reliable result.

Table 1 Results of the 1st survey (Contents of the 2nd survey)

No.	Contents
1	Is the interface design clean and good looking?
2	Does the design match?
3	Does icon design harmonize with reality?
4	Is the status of receiving and sending information clear?
5	Is the navigation clear?
6	Is important information clearly visible?
7	Can interface design prevent user error?
8	Is the information help holistic and effective?
9	Is it interactive and easy to use?

As a result of the second peer review, the app with the lowest score was removed in accordance with the criterion of maintaining at least one dynamic icon in each group. Tongcheng (199 points) was excluded from Group I, Youxiake (175 points) from Group II, and Ruotu (123 points) from Group III, resulting in the final sample of nine apps.

In the third expert review, the sample selection principle was established through rigorous research and discourse. The principle involved the following steps: Firstly, the sample should encompass both new and existing products. Secondly, a comparison and analysis was performed between apps with high ratings and those with low ratings. Thirdly, research subjects with distinct characteristics were selected. Fourthly, the sample in each group should comprise both apps with dynamic icons and those without dynamic icons in the same environment. The final six samples were selected based on this principle, as depicted in Tab. 2.

Table 2 Study Subject

APP Icon	English Name
	Ctrip
	Fliggy
	Mafengwo
	Lvmama
	Qyer
	Cuttlefish

3.1.2 Task Selection

The Task Test was implemented to assess the usability differences among the various functions of the travel app. The main functionalities of travel apps include transportation selection, travel planning, accommodation and situation sharing. In this study, a scenario-based task was designed to reflect the product positioning of travel apps and the role of dynamic icons. The scenario test was designed for a hypothetical free trip to Lijiang, China and consisted of the following tasks:

(1) Ticket reservation - Participants were required to complete a reservation form for a ticket, with no payment required.

(2) Travel planning - Participants were tasked with confirming a free travel itinerary, including attractions and courses.

(3) Accommodation reservation - Participants were instructed to reserve a room in Lijiang at a discounted rate, with convenient scheduling and no payment required after completing the business hotel reservation form.

(4) Situation sharing - Participants were required to share a travel diary by adding a random picture and the text "Hello, world".

3.2 Experimental Design

The objective of this study is to evaluate the user experience of selected travel applications by assessing both subjective and objective data based on task performance. The study focuses on examining the perceptual usability characteristics of subsystem navigation icons and analyzing the relationship between dynamic selection behavior and perceptual usability using objective indicators of dynamic icon usability. Additionally, the study aims to compare the overall system usability across various interface types of competing products under similar conditions.

Design and develop scenarios based on competitive experiments. The independent variables of the experiment were dynamic and static icons, while the dependent variable was usability. The experiment included six apps, of which three had dynamic interfaces with dynamic icons, and the other three had static interfaces without dynamic icons. The dynamic interface apps were considered as the dynamic group, while the static interface apps were considered as the static group. The six apps selected for the experiment were ranked based on their download popularity, with Ctrip and Fliggy in group I, Mafengwo and Lvmama in group II, and Qyer and Cuttlefish in group III. The letter "M" indicated the presence of a dynamic icon in the app. Specifically, the dynamic icon of Ctrip(M) was the navigation icon of task 3, the dynamic icon of Mafengwo(M) was the navigation icon of task 3, and the dynamic icon of Qyer(M) was the navigation icon of task 4.

3.2.1 Evaluation Indicators

In terms of evaluating usability, it is necessary to take into account the specific context of use and the individual's needs, preferences, and characteristics. A thorough analysis of usability in a system involves objective evaluation indices, such as the number of clicks or task completion time, and subjective evaluation indices, such as perceived ease of use and satisfaction. In addition, the application of dynamic icons in the subsystems can affect the perceived usability of the system. Therefore, evaluating the usability of the dynamic icons and subsystems in addition to the overall system is critical to comprehensively analyze the usability of the system.

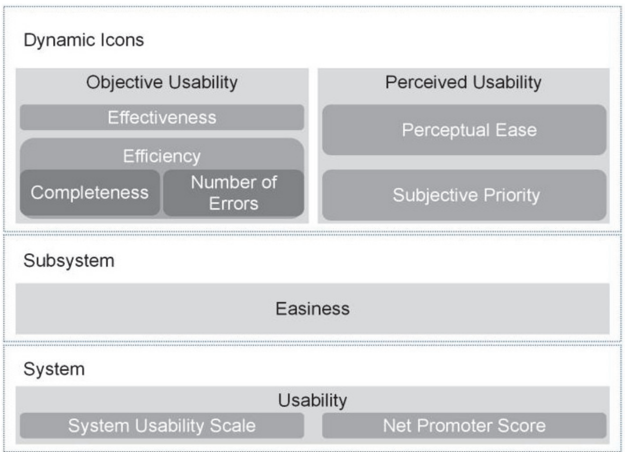

Figure 1 Evaluation Indicators

ISO9241/11 defines usability through three metrics: effectiveness, efficiency, and satisfaction. While effectiveness and efficiency are objective evaluation indices, satisfaction is a subjective indicator. The navigation gateway of subsystems, which is the application of dynamic icons, can impact the ease of use of the subsystems and, as a result, influence the overall usability of the system. Therefore, as depicted in Fig. 1, evaluations must be performed from three perspectives, which include dynamic icons, subsystems, and the system. Perceived usability is evaluated through subjective evaluation indices, and a comparison with objective usability evaluation results can provide a more comprehensive analysis of the usability features of dynamic icons.

Subjective evaluation indices are commonly used in the assessment of perceived usability. These indices include the System Usability Scale (SUS), the Usefulness, Satisfaction and Ease of use (USE) questionnaire, and the Perceived Ease of Use (PEOU) and Perceived Usefulness (PU) scales. In addition, objective evaluation indices, such as task completion time and error rates, can also be used in evaluating the usability of the system.

In this study, objective usability indicators are derived from experimental records, while perceptual usability and system usability indicators are obtained from evaluation questionnaires. Objective usability indicators include effectiveness and efficiency. Efficiency is assessed by completion time, which comprises two-level indicators of completeness and the number of errors. Completeness refers to the proportion of users who completed the task among all subjects. Completion time is the time taken by the user to complete one task. The number of errors is the number of instances in which the user failed to select the experimental icon while performing a given task. Completion time is measured as the time taken by the user to select the subsystem task icon.

The perceptual usability index consists of the perceived ease and the subjective priority. The Customer Effort Score (CES) is used to evaluate the ease of using subsystems, and the fourth question is used to evaluate the perceptual ease of icons. System usability is assessed using the System Usability Scale (SUS) and the Net Promoter Score (NPS). CES measures the difficulty users face when solving problems using the app, while subjective priority is used to identify the disturbing features of the interface that affect usability. Due to the complexity of operating technical products, such as apps, users demand reduced barriers to entry into technology. By comparing and utilizing the differences between dynamic and static icons in competing apps, it has been observed that dynamic icons can lower the difficulty of interface navigation. To this end, CES has been designed to eliminate or reduce obstacles faced by users during app usage, and to reflect the impact of dynamic icons on subsystem usability. Therefore, it has been applied to evaluate subsystem ease-of-use. SUS measures the overall perceptual usability of the app after performing a series of tasks. The evaluation based on SUS can reflect the effect of the dynamic icon on the system from the perspective of the ease of learning the system by deriving the learnability result. The NPS is a commonly used customer loyalty analysis indicator due to its simplicity and ability to provide a wealth of information. It is measured by asking users if they would recommend the company or product to a friend, which reflects their level of satisfaction and willingness to promote it.

In conclusion, the assessment of the usability must include the evaluation of dynamic icons, subsystems, and the overall system. Perceived usability must be assessed through subjective evaluation indices, and a comparison with objective usability evaluation results can provide a more comprehensive analysis of the usability features of dynamic icons.

3.2.2 Participants

Nielsen and Landauer conducted experiments and found that a usability test involving 15 participants could reveal 70% to 100% of usability problems [11]. Zapata suggested that the number of users required to evaluate mobile health apps ranges from 4 to 194, and that usability can be evaluated by less than 40 users for almost all apps [12]. To increase the objectivity and accuracy of experimental results, this study recruited 53 participants who are target users of the app. The study targeted nationwide users without any special restrictions on scope. The participants were aged in their twenties and were potential and real users of the app. Their average age was 21 ± 0.6 years, and the male-to-female ratio was 29:24. All participants were right-handed and had an average corrected visual acuity of 1.5 for the left eye and 1.6 for the right eye. They did not have color vision abnormalities or any other visual impairments that would affect their evaluation. All participants were frequent users of smartphones and were familiar with operating and interacting with them.

3.2.3 Experimental Procedure

The experiments were carried out on smartphones, specifically on the iPhone and Android phones. The choice of the operating system was based on familiarity, and the participants were asked to choose between iOS and Android. Similarly, the keyboard input method was based on familiarity, and participants were asked to select between Cheonjiin and QWERTY keyboards. The tests were conducted using a stopwatch and recorded using a cell phone camera. A total of six apps were tested, with four tasks for each app. The test was divided into three stages: pre-test preparation, test progress, and post-test evaluation.

In the pre-test preparation stage, participants were provided with a pre-test questionnaire (see appendix) and an experiment manual (see appendix) to familiarize themselves with the test environment, tasks, and installation of the test app. The aim was to reduce the variables associated with an unfamiliar experimental environment and create a test environment that is similar to the actual usage situation. All tasks lasted no longer than five minutes, and if the time exceeded five minutes, the task was considered a failure. During the task, participants were asked to describe their experience using the app and any interaction inconvenience through thought dictation. This included interface design elements such as graphics, images, icons, colors, animations, and layouts, as well as interaction logic and task procedures. If a participant encountered any problem, they could ask the experimenter for help. Upon completion of all four tasks of each app, participants completed the assignment evaluation questionnaire (subjective priorities and CES). After completing all four tasks for each app, participants completed the evaluation questionnaires (SUS and NPS).

The test duration for each app was approximately 25 minutes, with a five-minute break after each test. Participants were required to test all six apps for approximately three hours. The entire test process was recorded and the experimenter recorded the task performance, including completeness, completion time, and number of errors.

4 EXPERIMENT RESULT

This study aims to evaluate the usability of a travel app for airline ticket reservation, travel schedule, room reservation, and situation sharing. Usability is evaluated through task performance and standardized questionnaires, which can scientifically quantify the usability of the app. In this experiment, the standardized questionnaires consist of a task evaluation questionnaire and an overall evaluation questionnaire.

The task evaluation questionnaire measures the user's sensory usability after completing each scenario task. It consists of two parts: subjective priority and the CES. The overall evaluation questionnaire measures the overall usability of the app system after performing all scenario tasks. It consists of two parts: the SUS and the NPS.

The data collected from objective and subjective indices are organized and analyzed in Excel. The t-test is performed using GraphPad 5.0 with a two-tailed analysis method. A p-value of less than 0.05 between the dynamic group and the static group indicates a significant statistical difference.

4.1 REVIEW THE USABILITY OF DYNAMIC ICONS
4.1.1 Objective Usability

(1) Effectiveness.

① Completion.

The effectiveness of the user's selection of dynamic icons increases with a higher degree of completion. The calculation method is the number of people who completed the task divided by the total number of people, multiplied by 100%. Tab. 3 illustrates the completeness of the tasks I, II, and III.

Table 3 Task Completion

	APP name	Task 1 flight booking	Task 2 travel plans	Task 3 Room Reservation	Task 4 Moment Share
I	Ctrip (M)	100	100	100	100
	Fliggy	100	100	100	83.02
II	Mafengwo (M)	100	100	100	100
	Lvmama	100	100	100	100
III	Qyer (M)	88.7	94.3	95.6	75.5
	Cuttlefish	49.1	88.7	86.8	100

Tab. 3 shows that, among the high levels of completion of tasks I and II, Fliggy's task 40 did not achieve 100% completion. Conversely, task III exhibited low completeness levels overall, except for the Cuttlefish task that reached 40% completion. Completeness had no significant correlation with dynamic icons, but had a positive correlation with app preference. This indicates the necessity of the stratified extraction method. Completeness serves as the basis for other indicators, where test data of users who failed the task are removed.

② Number of Errors.

The effectiveness of dynamic icons increases with a lower average number of errors committed by users when performing a task. Tab. 4 to Tab. 7 display the experimental results.

Table 4 Number of Errors in Task 1

	APP name	Mean ± SEM / times	*p* value
I	Ctrip (M)	0.19 ± 0.1	0.621
	Fliggy	0.25 ± 0.1	
II	Mafengwo (M)	0.37 ± 0.1	0.741
	Lvmama	0.42 ± 0.2	
III	Qyer (M)	0.98 ± 0.2	0.212
	Cuttlefish	1.50 ± 0.4	

In Task 1, the average score for users to book air tickets using apps I, II, and III is similar, with no statistically significant difference in the number of errors between the three groups.

Table 5 Number of Errors in Task 2

	APP name	Mean ± SEM / times	*p* value
I	Ctrip (M)	0.17 ± 0.1	1.000
	Fliggy	0.17 ± 0.1	
II	Mafengwo (M)	0.77 ± 0.4	0.508
	Lvmama	0.51 ± 0.2	
III	Qyer (M)	0.82 ± 0.2	0.497
	Cuttlefish	1.06 ± 0.3	

In Task 2, no statistically significant difference exists in the number of errors that occur when the user makes a travel plan using the apps I, II, and III because the average number of errors is similar.

Table 6 Number of Errors in Task 3

	APP name	Mean ± SEM / times	*p* value
I	Ctrip (M)	0.23 ± 0.1	0.543
	Fliggy	0.13 ± 0.1	
II	Mafengwo (M)	0.09 ± 0.1	0.678
	Lvmama	0.13 ± 0.1	
III	Qyer (M)	0.19 ± 0.1	0.133
	Cuttlefish	0.43 ± 0.1	

In Task 3, the average number of errors that occur when a user makes a ticket reservation using the apps I, II, and III is similar, with no statistically significant difference in the number of errors between the three groups.

Table 7 Number of Errors in Task 4

	app name	Mean ± SEM / times	*p* value
I	Ctrip (M)	0.15 ± 0.1	0.034
	Fliggy	1.57 ± 0.1	
II	Mafengwo (M)	0.09 ± 0.0	0.077
	Lvmama	0.26 ± 0.1	
III	Qyer (M)	0.35 ± 0.2	0.148
	Cuttlefish	0.09 ± 0.1	

In Task 4, the average number of errors that occur when users share situations using the apps of II and III is similar. Conversely, there is a statistically significant difference in I (p = 0.034). In I, the average number of errors in the dynamic group is the lowest. This indicates that the user interaction effectiveness of the dynamic group is higher in I. No significant difference in the number of errors exists in the other two groups.

Upon comparing the experiments of the four tasks, the number of errors was 43.48% higher in the dynamic group than in the static group when the experimental condition of the dynamic icon was the task navigation icon. Moreover, the number of errors was 44.44% less in the II dynamic group than in the static group, and the number of errors in the III dynamic group was 44.44% less than that of the static group. Finally, the number of errors in the dynamic group was 74.29% higher than that of the static group.

(2) Efficiency.

A shorter average completion time for a task indicates faster selection and clicking of task navigation icons, which in turn makes them more accessible to users. The experimental results, presented in Tab. 8 to Tab. 11, show the impact of task navigation icon types on completion times for four tasks.

Table 8 Completion Time of Task 1

	APP name	Mean ± SEM / s	*p* value
I	Ctrip (M)	3.47 ± 0.3	0.0043
	Fliggy	3.84 ± 0.31	
II	Mafengwo (M)	4.33 ± 0.4	0.534
	Lvmama	4.67 ± 0.4	
III	Qyer (M)	8.54 ± 0.7	0.378
	Cuttlefish	10.39 ± 0.7	

For Task 1, statistical analysis indicates that the time taken for users to select icon I (p = 0.0043) was significantly different from the time taken to select II and III. The dynamic group displayed a shorter average completion time across I, II, and III, with users selecting the dynamic icon faster. These findings suggest that task navigation icons in the dynamic group are more accessible to users.

Table 9 Completion Time of Task 2

	APP name	Mean ± SEM / s	*p* value
I	Ctrip (M)	2.65 ± 0.2	0.0009
	Fliggy	3.68 ± 0.3	
II	Mafengwo (M)	2.99 ± 0.3	0.008
	Lvmama	4.98 ± 0.5	
III	Qyer (M)	9.24 ± 0.8	0.545
	Cuttlefish	11.09 ± 0.5	

In Task 2, there was a statistical difference between I (p = 0.0009) and II (p = 0.008). The shorter average completion time of dynamic group I and static groups II and III indicates that task navigation icons in dynamic group I and static group III are more accessible to users.

Table 10 Completion Time of Task 3

	APP name	Mean ± SEM /s	*p* value
I	Ctrip (M)	2.65 ± 0.2	< 0.0001
	Fliggy	3.68 ± 0.3	
II	Mafengwo (M)	2.99 ± 0.3	< 0.0001
	Lvmama	4.98 ± 0.5	
III	Qyer (M)	9.24 ± 0.8	0.0005
	Cuttlefish	11.09 ±0.5	

For Task 3, statistical analysis shows that I ($p < 0.0001$), II ($p < 0.0001$), and III (p = 0.0005) were statistically different, with I and II showing a significant difference. The dynamic group displayed a shorter average completion time and users selected the dynamic icon faster. These results indicate that the task navigation icon in the dynamic group is more accessible to the user.

Table 11 Completion Time of Task 4

	APP name	Mean ±SEM / s	*p* value
I	Ctrip (M)	2.65 ± 0.2	< 0.0001
	Fliggy	3.68 ± 0.3	
II	Mafengwo (M)	2.99 ± 0.3	< 0.0001
	Lvmama	4.98 ± 0.5	
III	Qyer (M)	9.24 ± 0.8	0.0005
	Cuttlefish	11.09 ± 0.5	

In Task 4, there was no statistically significant difference in the time taken by users to select and share a dynamic navigation icon. The average completion time for the dynamic group across I, II, and III was shorter, with users selecting the dynamic icon faster. These findings suggest that the task navigation icon of the dynamic group is more accessible to the user.

Overall, the experimental results indicate that when the experimental condition for the task navigation icon was dynamic, the completion time for the dynamic groups was significantly faster than for the static groups. Specifically, the completion time of dynamic group I was 28.00% faster than that of the static group, dynamic group II was 60.04% faster than that of the static group, and dynamic group III was 26.92% faster than the static group. These results suggest that the task navigation icons of the dynamic group are more efficient and accessible to users.

4.1.2 Perceptual Usability

(1) Ease of Subsystem.

The objective of the CES indicator is to identify and eliminate or minimize elements in the service that customers value too much. This approach permits us to identify and remove any distractions that affect the service experience. However, since researchers are unable to know which obstacles interfere with improving usability, it is essential to differentiate barriers through the subjective priority of interface elements. In this experiment, we offered users the option to choose among six interface elements, including interface layout, shape, motion, color, image, and text. Given the complex system design of the app, it is necessary to measure the level of effort for each task (Tab. 12 to Tab. 15).

Table 12 CES in Task 1

	APP name	Mean ± SEM / points	*p* value
I	Ctrip (M)	6.34 ± 0.2	0.889
	Fliggy	6.30 ± 0.2	
II	Mafengwo (M)	5.60 ± 0.2	0.257
	Lvmama	5.90 ± 0.2	
III	Qyer (M)	5.19 ± 0.2	0.0002
	Cuttlefish	3.78 ± 0.3	

In Task 1, the average score for booking air tickets using apps I and II was similar. In contrast, III ($p = 0.0002$) showed a statistically significant difference as the mean score of the dynamic group was higher. In III, the interaction design difficulty of the dynamic group was lower. In the other two groups, the interaction design difficulty of the dynamic and static groups was similar. The average score was high, and the difficulty of the interaction design was low.

Table 13 CES in Task 2

	APP name	Mean ± SEM / points	*p* value
I	Ctrip (M)	5.94 ± 0.2	0.522
	Fliggy	5.78 ± 0.2	
II	Mafengwo (M)	5.98 ± 0.2	0.203
	Lvmama	5.66 ± 0.2	
III	Qyer (M)	5.53 ± 0.2	0.014
	Cuttlefish	4.77 ± 0.2	

In Task 2, the average score for planning a trip using the apps of I and II was similar. However, there was a statistically significant difference in III ($p = 0.014$). The dynamic group of III had a higher average score, which means that the difficulty of interaction design of the dynamic group in III was lower. For the measurement results of the other two groups, the difficulty of interaction design was low, as in Task 1.

Table 14 CES n Task 3

	APP name	Mean ± SEM / points	*p* value
I	Ctrip (M)	6.30 ± 0.2	0.655
	Fliggy	6.40 ± 0.1	
II	Mafengwo (M)	6.20 ± 0.1	0.375
	Lvmama	6.02 ± 0.2	
III	Qyer (M)	5.62 ± 0.2	0.098
	Cuttlefish	5.08 ± 0.3	

In Task 3, the average score for booking a room using the apps of I, II, and III was similar. Overall, there was no significant difference between dynamic and static groups. However, III ($p = 0.098$) was statistically similar. In III, the high average score of the dynamic group meant that the interaction design difficulty of the dynamic group was lower. The average score of the other two groups was higher than III, and the difficulty of interaction design was similar.

Table 15 CES in Task 4

	APP name	Mean ± SEM / points	*p* value
I	Ctrip (M)	6.00 ± 0.1	< 0.0001
	Fliggy	4.59 ± 0.3	
II	Mafengwo (M)	6.32 ± 0.1	0.032
	Lvmama	5.84 ± 0.2	
III	Qyer (M)	5.22 ± 0.3	0.014
	Cuttlefish	5.96 ± 0.1	

In Task 4, there was a statistically significant difference in the users' use of I ($p < 0.0001$), II ($p = 0.032$), and III ($p = 0.014$). The average score of the dynamic group of I, the dynamic group of II, and the static group of III was relatively high. Through this, it can be seen that the difficulty of the dynamic group interaction design of I and II was lower. The dynamic group mean score of III (5.22 ± 0.3) was relatively low, even within the same group. However, this number was higher than the average score of the static group of I (4.59 ± 0.3), which means that the dynamic group with low preference is more convenient to use than the static group with the highest preference.

To summarize the above, the average score of CES in I, II, and III showed an overall decreasing trend, and the difficulty of interaction design felt by users had a negative correlation with app preference. The CES of dynamic group I was 1.56% lower than that of a static group, the CES of dynamic group II was 2.90% higher than that of a static group, and the CES of dynamic group III was 12.40% lower than that of a static group.

(2) Perceptual Ease.

Table 16 Perceptual Ease of Task 1

	APP name	Mean ± SEM / points	*p* value
I	Ctrip (M)	6.43 ± 0.1	0.019
	Fliggy	6.32 ± 0.1	
II	Mafengwo (M)	5.81 ± 0.1	0.533
	Lvmama	6.09 ± 0.1	
III	Qyer (M)	5.02 ± 0.2	0.173
	Cuttlefish	3.60 ± 0.3	

The CES survey results for the fourth question are used to study the perceptual ease of navigation icons. We compare the evaluation results of dynamic icons and static icons to examine the perceptual ease of dynamic icons. Tab. 16 presents the evaluation results of the perceptual ease of navigation icons in Task 1.

We found a statistical difference in I of task 1 (p = 0.019). In I and III, the average score of navigation icon perceptual ease of the dynamic group was higher, while in II, the average score of the static group was higher. In other words, users perceive the navigation icons of I and III dynamic groups as easier to select, while for II static groups, they find the icons easier to select. Tab. 17 presents the evaluation results of the navigation icon perceptual ease of Task 2.

Table 17 Perceptual Ease of Task 2

	APP name	Mean ± SEM / points	p value
I	Ctrip (M)	6.25 ± 0.1	0.0009
	Fliggy	5.70 ± 0.2	
II	Mafengwo (M)	5.71 ± 0.2	0.904
	Lvmama	5.98 ± 0.2	
III	Qyer (M)	5.57 ± 0.2	0.056
	Cuttlefish	4.96 ± 0.3	

We found a statistical difference in I of Task 2 (p = 0.0009). In I and III, the average score of navigation icon perceptual ease of the dynamic group was higher, while in II, the average score of the static group was higher. This means that users find it easier to select the navigation icons of I and III dynamic groups, while for II static groups, they perceive the icons as easier to select. Tab. 18 presents the evaluation results of the navigation icon perceptual ease of Task 3.

Table 18 Perceptual Ease of Task 3

	APP name	Mean ± SEM / points	p value
I	Ctrip (M)	6.45 ± 0.1	0.242
	Fliggy	6.21 ± 0.1	
II	Mafengwo (M)	6.15 ± 0.2	0.772
	Lvmama	5.96 ± 0.1	
III	Qyer (M)	5.72 ± 0.2	0.202
	Cuttlefish	5.10 ± 0.2	

In Task 3, the average score of perceptual ease of navigation icons was higher in the I, II, and III dynamic groups. This indicates that users perceive the navigation icons of I, II, and III dynamic groups as easier to select. Tab. 19 presents the evaluation results of the navigation icon perceptual ease of Task 4.

Table 19 Perceptual Ease of Task 4

	APP name	Mean ± SEM / points	p value
I	Ctrip (M)	6.30 ± 0.1	<0.0001
	Fliggy	4.19 ± 0.3	
II	Mafengwo (M)	4.71 ± 0.3	<0.0001
	Lvmama	5.87 ± 0.1	
III	Qyer (M)	5.00 ± 0.3	0.004
	Cuttlefish	5.76 ± 0.2	

In Task 4, we found a statistically significant difference between I ($p < 0.0001$) and II ($p < 0.0001$), and a statistical difference in III ($p = 0.004$). The average score of navigation icon perceptual ease of the I dynamic group and the II and III static groups was relatively high. This means that users find it easier to select the navigation icons of the I dynamic group and the II and III static groups.

Based on a comprehensive comparison of the experimental results of the four tasks, we found that the perceptual ease of the I dynamic icon was 3.72% higher than that of the static icon, the perceptual ease of the II dynamic icon was 3.09% higher than that of the static icon, and the perceptual ease of the III dynamic icon was 13.19% lower than that of the static icon.

4.2 Consideration of System Usability

Once the subject completes the four tasks, the SUS is scored, and the results are presented in Tab. 20.

Table 20 SUS Results

	APP name	Mean ± SEM / points	p value
I	Ctrip (M)	69.5 ± 2.6	0.029
	Fliggy	62.2 ± 2.1	
II	Mafengwo (M)	62.7 ± 2.0	0.719
	Lvmama	61.7 ± 2.1	
III	Qyer (M)	57.6 ± 2.3	0.003
	Cuttlefish	47.5 ± 2.4	

To intuitively examine and interpret the meaning of the SUS score, Bangor, Kortum, and Miller established the relationship between letters, alphabets, acceptable range, and SUS score [13]. Additionally, the SUS score can be converted into a percentage grade, allowing for comparisons of the usability of measured products and systems with other products and systems in the overall database.

The SUS score for the Ctrip(M) is 69.5 points, with a percentage rating of approximately 57 points, indicating that it is more usable than about 56% of the same type of app. As demonstrated in Tab. 20, the average value of I is higher than that of II, and the average value of II is higher than that of III. SUS has a positive correlation with app download rankings preference. Furthermore, the score of all dynamic groups is higher than that of static groups, suggesting that the usability of the app system, including dynamic icons, is higher. Statistical analysis indicates a significant difference between I (p = 0.029) and III (p = 0.003), highlighting the impact of dynamic icons on the system usability of the app in both high and low preference situations.

4.3 Satisfaction

The NPS results for the app system are displayed in Fig. 2.

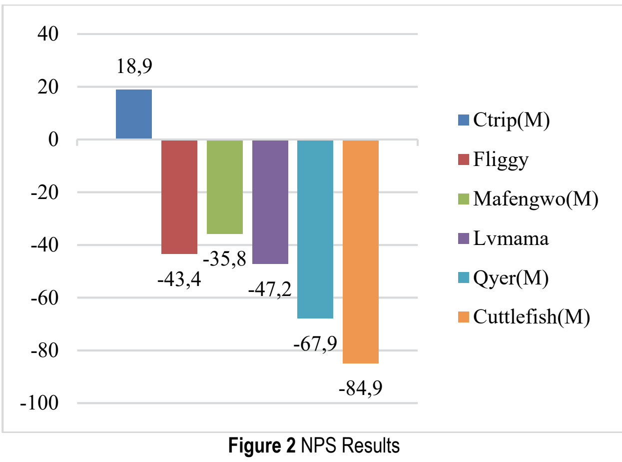

Figure 2 NPS Results

As shown in Fig. 2, the overall trend of NPS scores is almost identical to the app download ranking order. The NPS scores for all dynamic groups are higher than those for the static groups, indicating that the app system with dynamic icons is more likely to be recommended to users. Furthermore, the overall satisfaction with the app system, particularly those including dynamic icons, is higher.

5 CONCLUSION

The present study investigates the usability of dynamic icons and compares it to static icons, assessing both perceptual and objective usability. The evaluation of perceptual usability includes two indicators, namely perceptual ease and subjective priority. Results reveal that perceptual ease of dynamic icons is 3.72% higher than that of static icons in the case of I, 3.09% higher for II, and 13.19% lower for III. The subjective priority of dynamic icons is low, occupying the fifth position with 34.0% and 30.2% in Tasks 1 and 2, respectively, and the sixth position with 26.4% and 33.3% in Tasks 3 and 4, respectively, showing that users exert less force on dynamic icons. However, in conditions where only one dynamic icon is present in the mobile app interface, users exhibit strong attention to the dynamic icon, without perceiving it as a hindrance.

Objective usability is evaluated by assessing user behavior, specifically the ease of selecting the icons. Results show that the number of errors of the I dynamic icon is 43.48% higher than that of the static icon, while the number of errors for the II dynamic icon is 44.44% less and 74.29% higher for the III dynamic icon compared to the static icon. In terms of effectiveness, results differ depending on the test subject, but a significant difference is found between I and II ($p < 0.0001$). When a navigation icon is a dynamic icon in the task, completion time for the I dynamic group is 28.00% faster than the static group, 60.04% for the II dynamic group, and 26.92% for the III dynamic group, indicating that dynamic graphics induce users to select the dynamic icon more quickly, changing their navigation behavior.

Subsystems' ease of use is also evaluated, and results show that the dynamic icons of I and II have a minimal effect, while the III dynamic icon decreases subsystem ease by 12.40%. Synthetic analysis of thought dictation and user observations reveals that participants initially store information in short-term memory, concentrating on retrieving information related to the task. Dynamic icons stay longer in the visual attention phase when testing the task, making them less understood and remembered by users.

Finally, system usability and satisfaction are evaluated. Results show a statistical difference between I ($p = 0.029$) and III ($p = 0.003$), indicating that dynamic icons' attention-grabbing feature positively affects the system. Furthermore, excellent dynamic icon design and application can give an excellent expression even in apps with low preference, showing a positive effect on system satisfaction. Overall, while the perceptual usability of dynamic icons is not significantly different, the high objective usability of dynamic icons is evident, probably because their attention-grabbing feature induces stronger user attention.

Acknowledgments

The work was supported by the Social Science Foundation of Jiangsu Province (No.21YSD001).

6 REFERENCES

[1] Hyewon, Y. (2021) *Study on the effect of mobile app UI design factors on utilitarian value and continuous use: focusing on shopping app.* Doctoral dissertation, Hannam University.
[2] Lan, P. (2017). *Tourists Adoption Behavior on Mobile Tourism Application User experiment.* Doctoral dissertation, Xiamen University.
[3] Yang, T., Linder, J., & Bolchini, D. (2012). DEEP: design-oriented evaluation of perceived usability. *International Journal of Human-Computer Interaction*, *28*(5), 308-346. https://doi.org/10.1080/10447318.2011.586320
[4] McGee, M., Rich, A., & Dumas, J. (2004, September). Understanding the usability construct: User-perceived usability. *Proceedings ofthe Human Factors and Ergonomics Society Annual Meeting*, *48*(5), 907-911. https://doi.org/10.1177/154193120404800535
[5] Sauro, J. & Lewis, J. R. (2009, April). Correlations among prototypical usability metrics: evidence for the construct of usability. *Proceedings of the SIGCHI conference on human factors in computing systems*, 1609-1618. https://doi.org/10.1145/1518701.1518947
[6] Brooke, J. (2013). SUS: a retrospective. *Journal of usability studies*, *8*(2), 29-40.
[7] Lewis, J. R. (2018). Measuring perceived usability: The CSUQ, SUS, and UMUX. *International Journal of Human–Computer Interaction*, *34*(12), 1148-1156. https://doi.org/10.1080/10447318.2017.1418805
[8] Ardito, C., Costabile, M. F., Marsico, M. D., Lanzilotti, R., Levialdi, S., Roselli, T., & Rossano, V. (2006). An approach to usability evaluation of e-learning applications. *Universal access in the information society*, *4*(3), 270-283. https://doi.org/10.1007/s10209-005-0008-6
[9] Zhang, P., & Li, N. (2005). The importance of affective quality. *Communications of the ACM, 48*(9), 105-108. https://doi.org/10.1145/1081992.1081997
[10] https://www.iresearch.com.cn/Detail/report?id =3889&isfree=0
[11] Nielsen, J. & Landauer, T. K. (1993). A mathematical model of the finding of usability problems. *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, 206-213. https://doi.org/10.1145/169059.169166
[12] Zapata, B. C., Fernández-Alemán, J. L., Idri, A., & Toval, A. (2015). Empirical studies on usability of mHealth apps: a systematic literature review. *Journal of medical systems*, *39*(2), 1-19. https://doi.org/10.1007/s10916-014-0182-2
[13] Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies*, *4*(3), 114-123.

Contact information:

Yan LI, PhD
School of Textile Garment and Design, Changshu Institute of Technology,
No.99 South Third Ring Road (Donghu Campus),
Changshu 215500, Jiangsu, China
E-mail: yamtoon@foxmail.com

Yu Liang GUO, Associate Professor
(Corresponding author)
School of Textile Garment and Design, Changshu Institute of Technology,
No.99 South Third Ring Road (Donghu Campus),
Changshu 215500, Jiangsu, China
E-mail: 798255965@qq.com

Wei WEI, PhD
(Corresponding author)
School of Textile Garment and Design, Changshu Institute of Technology,
No.99 South Third Ring Road (Donghu Campus),
Changshu 215500, Jiangsu, China
E-mail: doublewei@cslg.edu.cn