

# Assignment 3: “Comparison of Expert and User Based Evaluation approaches in UI” Report

Eugene O. Casmin - 2528933 <sup>1</sup>

*Resumen*— As is the case with development cycles, the penultimate stage of UI design involves, of course, evaluating the designed user interface. There is a multitude of approaches that can be employed to achieve this purpose. The implementation of each of these approaches is dependent on a myriad of factors such as available financial and temporal resources, the type of target market for the UI, etc.

This paper explores UI evaluation of the IsBank mobile app in a bid to compare expert and user based evaluation techniques in UI evaluation.

*Palabras clave*— UI design, UI evaluation approaches, expert based evaluation, user based evaluation.

## I. INTRODUCTION

### A. The IsBank App

THE IsBank app to be used for this case study is your typical mobile banking application. It is easily the most widely used means of performing banking operations by the bank’s customers.

The application has provisions for the user to perform a variety of tasks of differing difficulty and sensitivity. Users can accomplish tasks as simple as opening new accounts (other than the primary account which is needed before one can access the mobile application; new accounts range from fixed term deposit accounts, investment accounts, etc) and buying and selling foreign currency to more sensitive task such as transferring money to other IsBank customers and even customers of other banks.

As we have seen throughout this course, the design of the user interface is integral to how effectively the users perform the tasks they hope to accomplish. The goal of evaluating the user interface is to verify whether the design and usability principles have been adhered to by the makers of the system.

In this report, we will be performing a comparison between expert based evaluation and user based evaluation approaches applied in evaluating the IsBank App.

## II. EXPERIMENTATION METHODOLOGY: EVALUATION ENVIRONMENT SETUP & EVALUATION

In this section, we will elaborate the process by which we set up the environment to perform the evaluation via the two approaches we aim to compare. The setup process involves selecting some fundamental tasks (for which we will explain why we chose the

tasks we chose), we will briefly introduce our test subjects (whose real identities shall remain anonymous throughout this report but shall be referred to with pseudo-names), we will continue to mention the evaluation techniques we have chosen to compare and finalise this section by describing any additional assumptions that we took into account for our evaluation environment.

### A. Fundamental Tasks

For our evaluation process, we selected two fundamental tasks which we have elaborated:

1. Accessing account information: The IsBank app, as any banking application ought to, allows the user to view the details of their accounts. Such details include the account balance in real time, their withdrawal history for a selected period of time as well as all their other accounts (besides their primary account, users may have foreign currency accounts, investment accounts, fixed deposit accounts and so on). Since this functionality is the one that we inferred from our test subjects to be the most utilised, we chose it as the “easy” task in our evaluation.
2. Transferring money to non-IsBank user: Additionally, the IsBank app provides the users with the ability to send as receive money to and from users of both IsBank and other banks. Since receiving money is a fairly straightforward procedure that just involves the user receiving a notification and checking their account balance (this is a testament to our assertion that the first task we chose is, indeed, fundamental as it is a composite task for other operations as well), we opted to select this task (transferring money to non-IsBank user) as our “difficult” task in our evaluation.

### B. Test subjects

In this section, we will introduce our test subjects and why elaborate the importance of each of them in our evaluation study.

Stratified sampling was used to select the test subjects with the label of the stratum being “expertise level” as explained below.

1. Test subject **A** (as they will be referred to throughout this study): Is a male first year student at METU NCC. The test subject is conversant with both Turkish and English languages to

<sup>1</sup>Computer Engineering Dept., METU NCC, e-mail: eugene.owilla@metu.edu.tr

a reasonable degree. The test subject has been using the IsBank mobile application for three months for the banking transactions being evaluated i.e., checking his account balance as well as sending and receiving money (as of the day of the study). However, he has not sent money directly to a non-IsBank user before. Thus, test subject A qualifies as our *novice* user/test subject for the user-based evaluation.

2. Test subject **B** (as they will be referred to throughout this study): Is a female second year aerospace engineering student at METU NCC. The test subject is reasonably conversant with Turkish language but is a native speaker of English. The test subject has been using the IsBank mobile application for two and a half years for the banking transactions being evaluated i.e., checking her account balance as well as sending and receiving money (as of the day of the study). However, she has not sent money directly to a non-IsBank user within the past nine months. Thus, test subject B qualifies as our *intermediate* user/test subject for the user-based evaluation.
3. Test subject **C** (as they will be referred to throughout this study): Is a male fourth year psychology student at METU NCC. The test subject is conversant with both Turkish and English. The test subject has been using the IsBank mobile application for five and a half years for the banking transactions being evaluated i.e., checking his account balance as well as sending and receiving money (as of the day of the study). The subject frequently sends and receives money to and from non-IsBank as well as IsBank users and he has sent money directly to a non-IsBank user within the past fortnight. Thus, test subject C qualifies as our *advanced* user/test subject for the user-based evaluation.
4. Test subject **E** (as they will be referred to throughout this study): Is a software engineering student. The test subject is a native English speaker. The test subject has never used the IsBank mobile application before but has worked in the development of and used other banking applications for the banking transactions being evaluated i.e., checking his account balance as well as sending and receiving money (as of the day of the study). The subject frequently sends and receives money to and from customers of the bank whose app he uses as well as non-customers and he has sent money directly to a non-customer within the past month. Thus, test subject E qualifies as our *expert* user/test subject for the expert-based evaluation.

The table below summarizes the user demographics.

	TS A	TS B	TS C	TS E
<b>IsBank app expertise</b>	Low	Medium	High	Low
<b>Banking apps expertise</b>	Low	Medium	High	High
<b>HCI knowledge</b>	Low	Low	Low	High
<b>Turkish proficiency</b>	Medium	Medium	High	Low
<b>English proficiency</b>	Medium	High	High	High

TABLE I  
USER DEMOGRAPHICS

### C. Evaluation Methods

In this subsection, we will introduce and elaborate on the evaluation methods that are the subject of comparison in this study. We will further explain the rationale behind the evaluation methods we have chosen and their suitability to our test environment.

1. In expert-based evaluation, a designer or HCI expert assesses a design based on known or standard cognitive principles or empirical results. Expert-based evaluation techniques are also referred to as expert analysis techniques. Examples of expert analysis methods include: heuristic evaluation, cognitive walk-through, and review-based evaluation. With this in mind, we chose a *heuristic analysis* as the method for expert-based evaluation. The rationale behind opting for this method lies in the need to use an approach that ensure we can ask the expert targeted design evaluation questions that can guide them to perform the evaluation within the shortest time possible. The rudimentary questionnaire employed involves the ten design heuristics by Nielsen Norman Group from [2] combined with a Likert scale for each heuristic on which the expert is expected to fill in their degree of agreement as to whether the corresponding design heuristic was adhered to by the IsBank application throughout provision of the service they were attempting to perform. As is a requirement of heuristic analysis, the expert then compiles a report outlining why they evaluated the UI as having met the design heuristic or not. We would like to reiterate that the questionnaire is only a guidance tool to distill the expert's evaluation and not the primary findings of the expert's evaluation.

The assertions which the expert was asked whether or not they agree with were framed as follows:

- Visibility of system status: The operations needed to accomplish the task were clearly

visible and accessible every step of the way.

- Aesthetics and minimalism: The application presents UIs in an aesthetically pleasing way and the UI's degree of minimalism is optimum. The information provided on the UI is focused on the essential and is enough to guide the user without needing additional explanation but not too much to crowd the interface or overwhelm the user.
  - User control and freedom: The application provides the user the chance to safely abandon an incomplete operation in the event of a mis-step.
  - Consistency and standards: The operations on every available UI window throughout the task are consistent. There are no operations that produced conflicting outcomes. Performing the same task on different platforms entails the same set of operations.
  - Error prevention: The UI notified the user of a possible error and allowed the user the chance to correct the error before proceeding.
  - Recognition rather than recall: The tasks on the UI are enough to guide the user to perform any operation they would like to. The user need not depend on their memory of previous use of the system or a similar system.
  - Flexibility and efficiency of use: The UI provides the user the chance to customise the view. The user can easily and safely select a UI appearance, language or other user-preferred functionality using straightforward operations.
  - Match between the system and the real world: The user is able to employ knowledge of daily life tasks/objects to perform a task or operation or interpret an otherwise unfamiliar task without producing an undo-able or unpredictable effect.
  - Help user to recognise, diagnose and recover from errors: The UI notified the user of a possible error and allowed the user the chance to correct the error before proceeding. The UI allowed the user the chance to undo an erroneous operation.
  - Help and documentation: The UI provides adequate, easily comprehensible documentation which assisted the user upon implementation. The five-point (ranging from 'Strongly Disagree' to 'Strongly Agree') Likert scale was then provided underneath each of the assertions for the evaluation of each fundamental task.
2. In user-based evaluation, we also used a **questionnaire**. However, since our goal is to get the *users' opinion*, the questionnaire provided to the test subjects was open-ended, shorter than the one elaborated for the expert-based evaluation and a lot more inclined toward getting a measure of the users' experience. For this reason, quantitative questions were avoided in a bid to avoid the users distilling answers thus leading to

missing UX information.

The questionnaire was administered in-person after observing the users interact with the application (although we did not interfere with the interaction process because this would change our evaluation to walk through methods).

The questions provided were as follows:

- How would you describe your emotion/feeling/mood while performing the tasks in one or two words? (To get information about the emotion elicited when using the app).
- Would you classify this as a good UI? Briefly state Why or why not? (To get a general comment about what counts as a "good UI" to them)
- How would you improve it in one or two operations? (To give them the chance to make changes that would most improve UX in their opinion. Also a follow up on question 2.)
- Would you recommend this banking application to your loved ones? (The general sentiment is that people would give a more well thought out answer if it, even hypothetically, involves an individual whose welfare they are mindful of).

### III. FINDINGS

In this section, we will analyse the results of the expert and user based evaluation performed for the IsBank mobile app for the two chosen tasks.

We will provide the results grouped by the evaluation approach (either expert or user based) and proceed to compare and contrast the findings.

#### A. Expert Based Evaluation

The findings of the expert-based approach are as shown below.

On the five-point Likert scale, the key is as follows: Scores [1 - 'Strongly Disagree', 2 - 'Disagree', 3 - 'Neutral', 4 - 'Agree', 5 - 'Strongly Agree']. It is important to note that in this case, the neutral value was primarily used for heuristics that could not be tested for in the corresponding task.

	Task 1	Task 2
<b>Visibility</b>	5	5
<b>Aesthetics &amp; Minimalism</b>	5	5
<b>User control and freedom</b>	5	5
<b>Consistency &amp; standards</b>	5	4
<b>Error prevention</b>	3	4
<b>Recognition vs. recall</b>	5	4
<b>Flexibility &amp; efficiency</b>	5	4
<b>Mapping</b>	3	5
<b>Error recognition, diagnosis &amp; recovery</b>	5	5
<b>Help &amp; documentation</b>	5	4

TABLE II  
FINDINGS FROM EXPERT-BASED EVALUATION

From the table, we can see that the expert user found it hard to perform the second task in a manner that was viable for him.

#### A.1 Expert's heuristic evaluation report

This report entails the experts detailed comments explaining his decisions on whether or not the ten Nielsen Normann design heuristics were met.

1. **Visibility of system status:** The UI adheres to the visibility heuristic. During both tasks/operations, the state of the system is conveyed to its users to a reasonable degree. The system keeps the user informed on what is going on through appropriate feedback within a reasonable time to empower the user to make decisions based in this information. The amount of information conveyed is not too much to distract the user and is just enough for the user to act on. The UI therefore achieves open and

continuous communication with the user. A notable case of this is, for instance, when the user chooses the option to "View all my accounts", a spinning wheel appears to communicate to the user that the requested screen is loading. Similarly, any option chosen along the path of the second task elicits the spinning wheel to communicate to the user that their request is being handled and they will not keep tapping on the screen.

2. **Match between the system and the real world:** The system speaks the user's language; with words, phrases and concepts familiar to the user rather than system oriented terms. The UI follows real-world conventions and the events constituting [the chosen] operations and subsequent information appears in a logical order. The system uses direct plain language in availing the options i.e., "View all my accounts" and the option that the users pick to transfer money to a non-IsBank user labelled "To Another Bank Account" are good examples of plain language. Additionally, the sequence of operations that the second task entails, follow a logical order. For instance, the user is asked to choose whether they'd like to send money to an IsBank user or a non-IsBank user, then they are asked whether the recipient account is already among their registered recipient.
3. **User control and freedom:** For the chosen tasks, the system UI provides clearly marked emergency exit for when the user performs an unwanted operation. The UI supports user undo and redo operations in line with this heuristic. This is exhibited by the undo (back arrow) button availed on the screen at each step of the second task to allow the user to leave an unwanted state in the event that they select a wrong option as well as the redo (front arrow) button provided if the user would like to revert to the state they left, for instance.
4. **Consistency and standards:** The UI is predictable and learn-able. The user need not wonder about conflicting outcomes for the same set of actions. Tapping on an option on the menu sends you to a sub-menu or opens up the interface to perform the operation you chose to do. This is maintained throughout the UI. This is in line with the expectations for internal consistency. Accessing the UI on iPad to perform the chosen operations requires the user to perform the same operations as on the iPhone. This is in line with the expectations for external consistency.
5. **Error prevention:** The application provides adequate support for error prevention. For example, the user is prompted to confirm the entered recipient's details in the second task. The application provides an unobstructed view of all the details before asking the user to confirm the transfer of funds.

6. Recognition rather than recall: In both tasks, the success of the operations is clearly more dependent on the user recognising the label on the menu provided. With the exception of the inference needed to make the mapping between the “Foreign Currency Transactions” label and transferring money to non-Turkish banks, the application reduces the cognitive load of the user to a reasonable degree.
7. Flexibility and efficiency of use: The application provides adequate support for the user to perform simple but crucial operations such as copying and pasting swift codes and bank account numbers to the UI which can be edited further. These operations are the most error prone when one is sending money to an unregistered account. Provision for such tasks improves user experience by providing flexibility and sharpening efficiency during use.
8. Aesthetics and minimalism: The UI design is focused on the essentials. The information provided on the UI is focused on the essential and is enough to guide the user without needing additional explanation but not too much to crowd the interface or overwhelm the user. The labels are clear and concise.
9. Help Users Recognize, Diagnose and Recover from Errors: The UI clearly informs the user when error has occurred by providing clear and concise error messages. The UI notifies the user what went wrong. The error messages provided have enough information and in a language that the user can understand, to notify the user why they are receiving the message. The UI offers a way to fix the error. The user is provided with placeholders that guide them on what to enter at each point when filling the recipient form in the second task. The UI also provides an example of a telephone number input in the event that the user has entered the wrong format.
10. Help and documentation: For the given tasks, it is reasonably easy to search for help on the UI. The documentation provided is focused on the user’s task. The documentation provided lists concrete steps to be carried out by the user. The language used is plain and not reliant on domain jargon.

### B. User Based Evaluation

The summarised findings of the user-based evaluation are shown in the table below.

	<b>TS A</b>	<b>TS B</b>	<b>TS C</b>
<b>Q1 - emotion</b>	Frustrated	Curious	Calm
<b>Q2 - good/bad UI</b>	Good	Good	Good
<b>Q3 - improvements</b>	clearer support for non-Turkish banks	clearer support for non-Turkish banks	provide pictorial tutorial
<b>Q4 - recommendation</b>	No	Yes	Yes

TABLE III  
FINDINGS FROM USER-BASED EVALUATION

Note: the app does provide support for transferring funds to non-Turkish banks but this requires selecting the “Foreign Currency Transactions” option.

As you would recall, the test subjects A, B and C are named in order of increasing longevity of use (NOT to be confused with the expert test subject D). The findings of the questions the test subjects were subjected to do indeed show this general trend.

Though not a recorded part of our methodology/results, it would be prudent to note that test subject A, for instance did exhibit a facial expression and demeanor to indicate frustration while evaluating the app. This could be seen by his stiffness in movement in posture and repeated attempts at trying to calm down. Additionally, the test subject had to restart the operation up to four times; further indication of anxiety-inducement that is characteristic of a novice user. It is therefore no wonder that the test subject chose not to recommend the application to a loved one as shown in the table above.

Consequently, test subjects B and C who were more familiar with the mobile application exhibited more confidence while handling the operations. Test subject B only had to restart the operation once but had to carefully read through all the options in the provided menus; which would be relatively time consuming in a real life scenario.

It was interesting to note that test subject C did not need to restart the operation but rather glided through the tasks constituting the operation, even asking the tester guiding questions such as whether they knew the exact bank details of the intended recipient and so on.

### C. Comparison and contrast

A comparison of the findings from the two approaches provides us with information upon which we could predicate that the IsBank mobile application could do with some improvements in their

money transfer operations to non-IsBank persons if they intend to attract new customers or app users.

This conclusion is further supported by the fact that the new or unfamiliar users (whether expert or novice users) have some difficulty in performing this operation but perennial users of the app already have some element of predefined faith that “the feature must be there. I just need to find it”, which is antithetic to the study of good UI design.

#### IV. PROS AND CONS

We can further compare the two evaluation approaches in terms of higher level metrics and glean the pros and cons as follows:

1. Time: In a case such as this with a limited pool of test subjects and time constraints, expert based evaluation methods take longer than user based approaches. As is evident in this study, expert based approaches are more targeted and elaborate. Moreover, if one is using an expert (as in expert based approaches), one would want to get the most information out of the evaluation hence they are likely to ask more questions and request more articulate answers. However, one might argue that in a scenario with unlimited time and a larger pool of test subjects, it would take longer for a user-based approach to yield any results that can be comparable to an expert-based approach for the same study.
2. Reliability: The results of expert-based approaches bear more weight owing to the fact that they are given based on standards and guidelines unlike user based approaches that are biased by personal preferences and thus lack objectivity.
3. Cost: Depending on the type of medium used in the evaluation, costs may vary. For instance, if one intends to use a questionnaire, the size of the questionnaire comes into question since as we have seen, the expert based approach is likely to require a more elaborate questionnaire which means higher costs. However, if the user based approach is also using questionnaires (even if shorter), we would need a larger pool of test subjects thus raising the costs. One tie-breaker might be whether or not the expert(s) in the expert-based study are to be remunerated; this would significantly impact the cost of the expert based evaluation. The general inference for cost analysis is that, if the evaluation is not time-sensitive, it would be cheaper to employ a user-based approach.

#### REFERENCIAS

- [1] *Psychology of Design*, Claus-Christian Carbon. *Department of General Psychology and Methodology, University of Bamberg, Bamberg, Germany; Bamberg Graduate School of Affective and Cognitive Sciences (BaGrACS), Bamberg, Germany; Research Group EPAEG (Ergonomics, Psychological Aesthetics, Gestalt), Bamberg, Germany* Cambridge University Press, October 2019
- [2] *10 Usability Heuristics for User Interface Design*, <https://www.nngroup.com/articles/ten-usability-heuristics/>