A SAMPLING METHODOLOGY FOR USABILITY TESTING OF CONSUMER PRODUCTS CONSIDERING INDIVIDUAL DIFFERENCES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ALİ EMRE BERKMAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
INDUSTRIAL DESIGN

JUNE 2010

Approval of the thesis:

A SAMPLING METHODOLOGY FOR USABILITY TESTING OF CONSUMER PRODUCTS CONSIDERING INDIVIDUAL DIFFERENCES

submitted by **ALİ EMRE BERKMAN** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Industrial Design Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen ______________________
Dean, Graduate School of **Natural and Applied Sciences**

Assoc Prof. Dr. Gülay Hasdoğan ______________________
Head of Department, **Industrial Design**

Assoc. Prof. Dr. Çiğdem Erbuğ
Supervisor, **Industrial Design Dept., METU** ______________________

Examining Committee Members:

Assoc. Prof. Dr. Gülay Hasdoğan ______________________
Industrial Design Dept., METU

Assoc. Prof. Dr. Çiğdem Erbuğ ______________________
Industrial Design Dept., METU

Prof. Dr. Giray Berberoğlu ______________________
Secondary Science and Mathematics Education Dept., METU

Assoc. Prof. Dr. Mehmet Asatekin ______________________
Industrial Design Dept., Bahçeşehir University

Assoc. Prof. Dr. Tayyar Şen ______________________
Industrial Engineering Dept., METU

Date: 24.06.2010

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : **Ali Emre BERKMAN**

Signature :

ABSTRACT

A SAMPLING METHODOLOGY FOR USABILITY TESTING OF CONSUMER PRODUCTS CONSIDERING INDIVIDUAL DIFFERENCES

Berkman, Ali Emre
Ph.D., Department of Industrial Design
Supervisor : Assoc. Prof. Dr. Çiğdem Erbuğ

June 2010, 388 pages

Aim of the study was to discuss and identify individual differences that influence the user performance during usability tests of consumer products that are known to prevent researchers to conduct systematic studies. The rationale behind the study was developing a tool for sampling in order to handle experiential factors as a variable rather than a source of error. The study made it possible to define and elaborate on constructs general interaction expertise (GIE) and general interaction self efficacy (GISE), and to devise a measurement scheme based on performance observation and attitude measurement. Both perspectives were evaluated with preliminary validity studies and it was possible to provide evidence on predictive validity of the tool developed. Furthermore, opportunities of utilizing the results in design and qualitative research settings were also explored.

Keywords: Usability testing, consumer products, general interaction expertise, general interaction self-efficacy

ÖZ

ÜRÜN KULLANILABİLİRLİĞİ TESTLERİNDE BİREYSEL FARKLILIKLARA DAYALI BİR ÖRNEKLEMLEME YÖNTEMİ

Berkman, Ali Emre
Doktora, Endüstri Ürünleri Tasarım Bölümü
Tez Yöneticisi : Doçent Dr. Çiğdem Erbuğ

Haziran 2010, 388 sayfa

Çalışma kullanılabilirlik testinde kullanıcı performansını etkileyerek yapılandırılmış araştırmaların gerçekleştirilmesini önleyen faktörlerin tanımlanmasını amaçlamaktadır. Temel amaç, bireysel farklılıklara dayalı örneklem oluşturmak için deneyim düzeyini bir hata kaynağı olmaktan çıkararak bir değişken olarak ele alınmasını sağlayacak bir araç geliştirmektir. Çalışma sonucunda genel etkileşim ekspertizi ve genel etkileşim öz yeterliği kavramları tanımlanmış, performans gözlemine ve tutum ölçümüne dayalı bir çoklu ölçüm yöntemi geliştirilmiştir. Geliştirilen yöntem kullanılabilirlik testleriyle beraber uygulanarak tahmin geçerliğine ilişkin kanıtlara ulaşılmıştır. Çalışmada elde edilen bulgular, ölçme perspektifinin ötesinde, tasarım ve niteliksel araştırma alanları bakımından da ele alınarak uygulama fırsatları araştırılmıştır.

Anahtar Kelimeler: Kullanılabilirlik testi, tüketici ürünleri, genel etkileşim ekspertizi, genel etkileşim öz yeterliği

To My Son Ozan

ACKNOWLEDGMENTS

It is a pleasure for me to thank many people who made this thesis possible:

It is difficult to overstate my gratitude to my advisor, Çiğdem Erbuğ. With her guidance, attention, encouragement and efforts to help me see the big picture and stay on tracks, she made this study possible;

To Thesis Supervising Committee Members and Thesis Examining Committee Members Mehmet Asatekin, Tayyar Şen and Gülay Hasdoğan for their comments and questions that broaden my perspective;

To Giray Berberoğlu for his inspirational advices and guidance;

To Bahar Şener, Evren Akar, Zeynep Karapars for their valuable input in scale development process and friendship;

To Pelin Atasoy for her friendship and encouragement throughout the process;

To Özgür Arun for his good ideas, endless support in analyses and friendship;

To my team of administrators Halil Karaçalı, Ezgi Nurali, Mazlum Akın, Özge Şaffak, and Ercan Tekdemir for their efforts and care in collecting reliable data;

To all the participants and respondents for their valuable input and patience;

To department secretaries Tülay Yıldız and Başak Timurçin for their positive attitudes and help in running the process smoothly;

To my family Necip Berkman, Ayşe Berkman Pierce for their patience and especially to my mother Gülden Berkman for her good advices;

Finally, but most entirely, to my wife Ceren for her endless patience, warm support and attention throughout this thesis. This study would be impossible without her.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

GIE : General Interaction Expertise

GIE_XEC : General Interaction Expertise Execution test that targets automatic behavior

GIE_PS : General Interaction Expertise Problem Solving test that targets controlled behavior

GISE : General Interaction Self Efficacy

GISE-S : General Interaction Self Efficacy Test

LEDQ : Learning Electronic Devices Questionnaire

NED : Number of Electronic Devices used

SEM : Structural Equation Modelling

UP : Usability performance

:

CHAPTER 1

1. INTRODUCTION

1.1. Rise of computer technology

After the developments in computer technology during 1970s and its rapid diffusion to various levels of society in the following years, the discipline of ergonomics, having gathered a vast body of knowledge in physical aspects of measurement and design in the past, had to rearrange itself according to the new circumstances. Helander (1997) states that the major shift of focus was from 'biological sciences' to mental issues, and owing to the extent of utilization of technology, to non-work activities as well. According to (Carroll, 2003), initial impetus for HCI was felt when linear design process adopted by software engineering, termed as waterfall development method, proved to be unsuccessful allocating 'software human factors' at the end of the process and software engineering found itself in the middle of a crisis. Although, ergonomics of programmer users was studied between 1960 and 1970, the problems of end-users was started to be recognized during 1970s (Smith, 1997). The most challenging issue faced with was the fact that the end-user audience of computer technologies was gradually being broadened. This process is schematized by Shackel & Richardson (1991) in four successive stages (see Table 1-1).

Table 1-1 Broadening audience of computer technologies

Computer type	Period	Users	Problems
Research machines	1950s	Scientists	• Reliability • All the programming is done by users
Mainframes	1960s – 1970s	Data-processing professionals	• Users of the output grow
Minicomputers	1970s	Engineers and other professionals	• Users still do programming • Usability becomes a problem
Microcomputers	1980s	Almost anyone	• Usability is the major problem

Note. Adapted from Human Factors for Informatics Usability by (Shackel & Richardson, 1991)

The increase of usability problems can be explained by the fact that the comparability between designer and users in terms of computer expertise, formerly avoiding serious problems to be encountered, was seriously disturbed after non-experts entered the scene.

The literature of ergonomics, indifferent to this upcoming issue at first, soon anticipated this prospective area with a rapid growth of interest (Meister, 1995).

According to Adler and Winograd (1992), although ergonomics was traditionally familiar to the issues of design of human – machine interface, the old approach had certain drawbacks as far the new problem domain is concerned. First, they argue that conventional models focused on lower levels of cognition such as sensation and perception, whereas new interaction required an understanding of complex functions. As a second argument, they emphasize the fact that modeling user as a system component was a narrow depiction, which makes it hard to grasp their active role. Thirdly, ergonomics was usually given a role of error reduction, where at a later stage of a development process the experts were asked to modify a given system in order to keep it within the limits and capabilities of users. Finally, the expert-centered evaluation methods that proved to be successful as far as physical capacities and low order cognitive facilities are taxed have lost their power within the hard-to-predict cases of complex interaction.

1.1.1. Diffusion of digital technologies

With the diffusion of digital technologies, problems that have been witnessed in the domain of personal computers (Shackel & Richardson, 1991) began to be observed in the use of once-humble products (Thimbleby, 1991). Together with this, conventional paradigm of consumer ergonomics was no more sufficient to embrace all the dimensions of user – product relationship.

Relatively complex cognitive processes that were in charge necessitated adoption of methods that traditionally belong to the domain of HCI. In a survey carried out in 1996, including 25 federated societies of IEA, 'usability of consumer products' was ranked as the third most important emerging area in ergonomics, leaving 'human computer interface' behind (Helander, 1997). Since 1990s, it is no more uncommon to come across with cases that consumer product are evaluated using techniques pertaining to HCI (e.g., Connell, Blanford, & Green, 2004; Garmer et al., 2002; Lauretta & Deffner, 1996).

Being a fundamental technique in HCI, usability testing is one of the most frequently applied techniques in both design and evaluation. As the observation of participant behavior forms the backbone of the technique, it is empirical and somewhat objective in character. Given this, usability testing is one of the most frequently resorted techniques when a systematic approach is required for eliminating evaluator biases as much as possible (Potosnak, 1988).

In the case of consumer products, while applying HCI-specific methods, adherence to conventions valid for HCI in a 'verbatim' fashion may cause incompatibilities. HCI theories and practice, 'user' is traditionally conceptualized as a professional, using a tool for sustaining her/his activity within the work domain. Therefore, the user profile exhibits a relatively homogenous profile.

Given these, for professional products, it is usually possible to determine the characteristics of target users and 'choose' the ones that represent the actual population as participants, with the help of observable attributes such as job experience, education, age etc.

In the case of consumer products, working on homogeneous 'subsets' is not plausible most of the time, given the fact that such products are usually intended for a larger portion of the population. Since anybody can be within the target profile, individual differences start to play an important role.

Diversity to be accommodated is quite large and many user characteristics, especially experiential ones, should be considered in order to ensure that design characteristics of the product being tested are reflected to results rather than individual differences. In the following chapters this will be discussed thoroughly.

1.2. Aim of the study

Aim of the study is to develop a framework to accommodate individual differences in usability tests and other user-centered design techniques in the case of consumer products, so that results are not affected by individual differences.

In order to accomplish this aim the following questions should be answered:

- What is the mainstream approach to sampling in usability studies?
- What are the individual differences that may affect usability test results? Do experiential factors play a significant role?
- How should experiential factors be approached so that they no more obscure link between design characteristics and usability performance?
- How can experiential factors be approached within a measurement perspective?
    - What may the manifestations of expertise be with digital products?
- How can this framework be utilized for evaluating design alternatives?
- How can this framework be utilized in qualitative research?

1.3. Structure of the thesis

In Chapter 2, the problem definition presented here will be discussed in detail by highlighting the problems with current approach to sampling and treatment of experiential variables as independent variables.

In Chapter 3, a construct definition and a model where experiential factors are defined with regards to what is acquired or retained will be discussed.

In Chapter 4, the prototypic tools developed to assess General Interaction Expertise, based on observation of the actual performance will be presented with relevant theory and empirical findings.

In Chapter 5, another assessment tool developed in order to assess another manifestation of GIE, namely General Interaction Self Efficacy will be discussed. Theoretical background and the development process will be presented in detail.

In Chapter 6, the findings of the empirical studies will be discussed in detail. Together with the nomothetic approach maintained throughout the study, other opportunities will be explored.

In the conclusion chapter the main outcomes and shortcomings will be discussed. The partial models utilized throughout the study will be presented as an integrated model, and finally future studies and opportunities for future work will be explored.

CHAPTER 2

2. DESIGN, USABILITY TESTING AND INDIVIDUAL DIFFERENCES

2.1. The link between design characteristics and usability

The rationale behind conducting a usability test is to measure (Nielsen, 1993) the high-level construct defined as ‘usability’ of a system, regardless of the organizational context in which it is conducted (Gray and Salzman, 1998). Therefore, as any other measurement instrument would claim to do, a usability test should be intended for its effectiveness to measure the targeted construct.

Regardless of the motivation behind testing a product, the aim is always to assess to what extent design is appropriate or the design decisions that may render a product inappropriate. In formative tests, products are tested during the development process in order to determine potential sources of usability problems and to generate design improvements so that the design is altered. Even in summative tests, products are tested so that designs may be assessed on their own or within a group of alternative/competing designs with regards to how usable they are. In each case the effect of design solutions on participants’ performance is being investigated, with the basic presumption that there is a causal relationship between them. In other words, when a product causes usability problems it is usually suggested that design has certain defects. The phenomenon pointed out by Norman (1988) that usability problems are mostly caused by the frequently coined "gap between designer and user" reflects a similar approach.

Therefore, it is not too much to suggest that the main motivation behind studying usability is to investigate the characteristics of the causal relationship between design and usability of a product.

In this regard, when a product does not seem to perform well in a usability test the cause of the misfit is expected to be design. All the other factors that may be in charge are regarded as nuisance variables and are tried to be eliminated.

The major disadvantage and the most powerful trait of the methodology of lab testing is regarded to be the reduction of real-life factors and isolating interaction in a controlled environment. The following lines by Woodworth that highlight why controlled conditions are crucial in inferential work opened up new opportunities in experimental research, and are worth quoting in full.

> An experimenter is said to control the conditions in which an event occurs. He [sic] has several advantages over an observer who simply follows the course of events without exercising any control.
>
> 1. The experimenter makes the events happen at a certain time and place and so is fully *prepared* to make an accurate observation.
>
> 2. Controlled conditions being *known* conditions, the experimenter can set up his experiment and repeat the observation; and, what is very important in view of

social nature of scientific investigation, he can report his conditions so that another experimenter can duplicate them and check the data.

3. The experimenter can systematically *vary* the conditions and note the concomitant variation in the results. If he follows the old standard "rule of one variable" he holds all the conditions constant except for one factor which is his "experimental factor" or his "independent variable." The observed effect is the "dependent variable" which is in a psychological experiment is some characteristic of behavior or reported experience. In an experiment on the effect of noise on mental work, noise is the independent variable controlled by the experimenter, and the dependent variable may be speed or accuracy of work or the subject's report of his feelings [...] With careful planning two or three independent variables can sometimes be handled in a single experiment [...] Whether one or more independent variables are used, it remains essential that all other conditions be constant. Otherwise you cannot connect the effect observed with any definitive cause.

(Woodworth, 1939; pp. 2-3)

Although such a methodological parsimony may not be required in the case of usability tests, the fact that one "cannot connect the effect observed with any definitive cause" if there are too many unknowns in the scene is a valid question directed towards usability tests of all sorts. In order to conduct analyses and draw valid conclusions, variables of concern should be somehow measured, even if the study is a non-experimental one (Spector, 1993).

According to the *classical test theory*, a measurement may not be freed of all its flaws and any act of measurement is subject to contamination, in terms of Spearman's *true score model* (1907; ctd. in. Spector, 1993).

$$\mathbf{X = t + e} \tag{1}$$

Where, **X** is the *observed value*, **t** is the *true score*, and **e** is the *error component*.

With an expansion of the *error component*, the conceptual formula can be stated as follows:

$$\mathbf{X = t + (e_r + e_s)} \qquad (2)$$

Where, $\mathbf{e_r}$ is the random error, and $\mathbf{e_s}$ stands for the *systematic error*. Whether a quantitative or a qualitative approach is adopted, the methodological challenge is to eliminate $\mathbf{e_s}$**,** and to reduce $\mathbf{e_r}$ by keeping with principles of good design and conduct, so that error component does not introduce a systematic bias, as far as the *observed score* is concerned (Cooper, 1998; Crocker & Algina, 1986).

In the case of usability tests many types of e_s may affect what was observed, despite the true fit between the design and the participant. A study that discusses the systematical error components in the case of usability testing was not located in the literature.

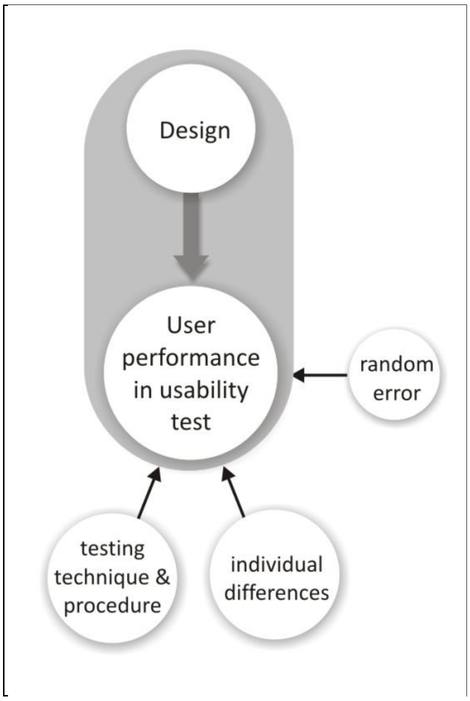


Figure 2-1 Possible factors that affect user performance in usability test

Testing technique and procedure may include mainly consistency problems, where every participant does not come across the same experience. For example, inconsistency in answering help requests and inadvertent questions directed to participants during a scenario may affect actual performance or the subject's feelings and ways of reporting them. Furthermore, the bugs and technical breakdowns witnessed during a test may also alter the results, so that some sessions may be lost entirely. Even a single hard-to-complete scenario skipped may alter the impressions about the product being tested and may affect a post-test satisfaction questionnaire to a great extent. Main texts on practical aspects of usability testing coves many of these as guidelines for testing (e.g see Nielsen, 1992; Dumas and Redish, 1993; and others)

Such errors may latently cause defying effects on test results and if are systematical in nature may ultimately alter the conclusions drawn. For example, suppose that a group of products are being tested and parallel sessions are necessary for methodological reasons or pure logistics. The style of administration exhibited by test administrators may deeply affect what was experienced and what was felt by the participants. Even, the gender and age of the administrator may induce a serious bias and a certain profile of participants may feel less anxious and more motivated during the test. Although such sources of error may cause serious problems, strictly followed procedures, technical competence, administrator training and consistency in administration may alleviate problems. Furthermore, it is possible to recognize such errors during the analysis phase.

Obscure sources of systematic error may not be recognized or located with such ease. Some types of individual differences among the participants may not be observed directly and may seriously obscure the causal link between design and usability. Observable or latent there are many types of individual differences that were treated as confounding variables in usability related studies.

2.2. Individual Differences and Usability

The branch of psychology studying differences among individuals is named as differential psychology. It is almost impossible to find a single aspect considering human beings where differences among individuals are so insignificant that they are easily neglected for the sake of parsimony (Carroll, 2003). Any user activity within an artificial system can be testified, without hesitation, to exhibit influences of individual differences in both quantitative and qualitative senses.

According to Cooper (1998) among the numerous merits of studying individual differences, four main reasons can be listed.

1. It is a challenging and intriguing issue of its own right.
2. Measurements of certain differences provide variables, thus increasing inferential accuracy and power of research.
3. Recognition of differences is useful and sometimes crucial in many practices—e.g. personnel selection, assessment of training, etc.
4. Individual differences can be investigated to predict behavior prior to performance.

Among the points listed above; 2 and 4 seem to overlap with the aims of this project.

2.3. Diversity of performance due to individual differences

Early studies that explored how HCI can benefit from differential psychology are reviewed and discussed in depth in an article by Egan (1988). Most of the early studies seem to concentrate on how general guidelines can be developed with an aim of accommodating individual differences in the design of systems for various tasks. The majority of research effort was to determine whether certain traits of individuals affect performance in common tasks carried out with computers such as information retrieval, text editing, accounting, and programming (e.g. Benbasat, Dexter and Masulis, 1981; Egan, Bowers and Gomez, 1982; Gomez et al., 1983; Vincente, Hayes and Willigies, 1987; Evans and Simkin, 1989; Nilsen et al., 1993). It should be noted that although such tasks were mostly carried out by a relatively homogenous user population, the ratio of best performance to the worst performance was found to be much higher than the typical ratios observed in conventional occupational settings. In order to grasp the significance of individual differences and the extent of diversity due to individual differences in observed measures of performance, Egan's seminal work (1988) is worth a concise review.

In his introductory lines, Egan states that there are three good reasons to approach to the issue of individual differences with a prescriptive approach rather than a descriptive one. First, he argues that it is common to observe performance differences as large as 20:1 for a particular task. What is surprising is that the differences can be explained by the diversity of users, regardless of the specific designs of the systems or training procedures. Egan identifies the number of errors made and time elapsed while recovering from errors as two main sources of performance differences in editing tasks. In accordance with this, he argues that tasks which do not tax cognitive resources or that are dominated by motor skills yield less difference in performance. Second, Egan states that as computer systems proliferate and are used by nonprofessional users as well, certain individuals will not be able to use such systems effectively, which may hinder success in the market. Lastly, it is argued that since these performance differences are not random they can be predicted and their causes can be identified for guiding better designs immune to individual differences (see Egan, 1988, p. 565 for a representation of the ideal system).

By reviewing a multitude of studies Egan concludes that causes of such variations in performance seem to be dominated by variables such as "experience, certain 'technical' aptitudes, age, and domain specific skills"(p. 552). *Experience*[1] was found to be usually the best predictor of performance if a group of users with varying levels of experiences are considered. However, it should be noted that the definition of experience adopted in these studies was quite problematical regarding how this attribute was represented (see Footnote 2, later to be discussed in this paper). *Technical aptitudes* that yield significant correlations with performance were identified as spatial abilities, reasoning and certain other aptitudes such as science / mathematics achievement. *Age* emerged to be a powerful predictor of learning performance if experience was controlled. In the case of text editing, after a brief period of learning, correlation between age and performance was observed to attenuate. *Domain specific skills* acquired with conventional tools were usually observed to hinder the performance with computerized tasks, since negative transfers were likely to occur and were more powerful as a domain specific skill become imbedded—i.e. as automatic processing is fully developed. Egan concluded that "domain specific knowledge begins to predict performance only after users have acquired some experience with the computer interface" (p. 557), in other words, after a certain level of *computer literacy* is acquired.

In a later study, by Dillon and Watson (1996), "over a century of work in differential and experimental psychology" (p. 631) was reviewed with an aim of enhancing user analyses typically carried out in HCI studies. The survey was

[1] Experience is usually conceived as pieces of information that consists of years-of-experience type data regarding a general or specific application domain—e.g. *no experience, two years of experience, more than three years of experience*, etc. The problems of such a definition will be later discussed in this article.

concluded with an inspiring discussion on ways in which the knowledge and research methods of differential psychology can be suitably added to the toolbox of HCI analyst. The relevant issues to be highlighted can be summarized as follows.

First, after years of research in psychometrics it was possible to identify a number of basic abilities; though, there are ongoing discussions about the relationships and the exact structure of high-order abilities (Cooper, 1998). Regardless of these meta-discussions, these basic abilities proved to be pragmatically useful in predicting performance regarding specific tasks. Second, design and analysis of systems can be improved with the knowledge accumulated. Such an improvement may open up the possibilities to generalize findings and to develop a data-driven user taxonomy, rather than pure arm-chair speculation. Third, certain individual differences such as reasoning and visual abilities can be associated with certain design characteristics of interfaces.

2.4. Current approach to sampling in usability tests

The literature of individual differences concerning usability seems to be restricted to professional and non-professional software domain. Studies that discuss individual differences in regards to consumer products with embedded software are rather scanty. The fact that individual differences regarding consumer products are much more significant in terms of all types of usability studies may be attributed to two main reasons. First, as interaction styles that could be exploited are increasing, designers started to assume more experience and ability on the user's side (Chen, Czerwinski and Macredie, 2000). Second, defining a clear-cut user population is quite difficult. In reality, ‘every person in the world’ can be a potential user for, say a cellular phone, produced by a multi-national company. Categories such as age, gender, education level or socio-economic status are far from having discriminatory power if compared to the attributes that directly influence performance (see Dunnette, 1976 for a full discussion), although some of such ‘generic’ categories may have a correlation with performance in some cases. Thus, a quite heterogeneous user population is confronted with, when one needs to conduct usability studies in the field of consumer products.

Causes and consequences of the heterogeneity of user population in the case of consumer products may best be illustrated with a speculative example:

Suppose that during the development process of an innovative cellular phone, the manufacturer wants to see whether users will easily adapt to the innovative interface. Furthermore, the manufacturer wants to compare the performance of this innovative design with its competitors and needs to verify that basic functions can be easily used by all users. Although, usability testing would be the right choice to fulfill those needs, results of the test would not be able to yield unambiguous results.

Firstly, the possibility that variance observed in user performance may be explained by individual differences causes methodological problems, and is hard to neglect especially in the case of consumer products. Some participants may not be able to complete even a single task successfully; interpretation of this result would really be trivial. Was it the interface’s design that caused too much problem for the participants? Was it the participants’ lack of experience with such innovative modes of interaction?

Secondly, when the task is to compare the design with its competitors a methodological problem with 'experiment design' arises. Suppose that interface (A) is decided to be compared with three other products (B, C and D). It is evident that a single test where each participant experiences all the interfaces is not possible, since such a test session would take too much time and it would be difficult to isolate and eliminate the effects of positive – negative transfer among interfaces. Therefore, one would look for experiment designs with more than one group. For example, there may be three groups where each competitor is compared with interface A, so that each participant uses only two interfaces instead of 4. In such a design, participants in each group should be comparable with regards to individual differences that may directly influence the test results.

Thirdly, the manufacturer in the example above would never know whether the sample was representative enough to infer that 'basic functions can be easily used by all users', regardless of the level of success observed in the tests.

The primary aim of any usability test should be to observe the effect of interface design on user performance, and eliminate all other interfering factors. Individual differences should be regarded as the most important factor to be eliminated or controlled since early studies show that huge variability in performance can be explained by individual differences among users, regardless of design or other factors (Egan, 1988). Experiential factors, among other individual differences, are known to have a significant effect on performance (e.g. Nielsen, 1993; Dumas and Redish, 1993).

Despite the famous phrase reminding participants that what is tested is the interface not their abilities, it is usually the participant's familiarity with digital interfaces that is being reflected in results.

2.5. When does heterogeneity really cause problems?

Although, the fact that experiential factors have a considerable effect on results indicates that a methodological flaw is present, this is not a criticism brought to the methodology of usability in general. Most of the time usability tests are conducted to uncover major problems and to have a rough idea about the fit between user and the system. It may be assumed that whether a test would be carried out in 'discount usability situations' (Nielsen, 1993) or for strict, inferential purposes (Potosnak, 1988) may determine how meticulously should external factors be controlled.

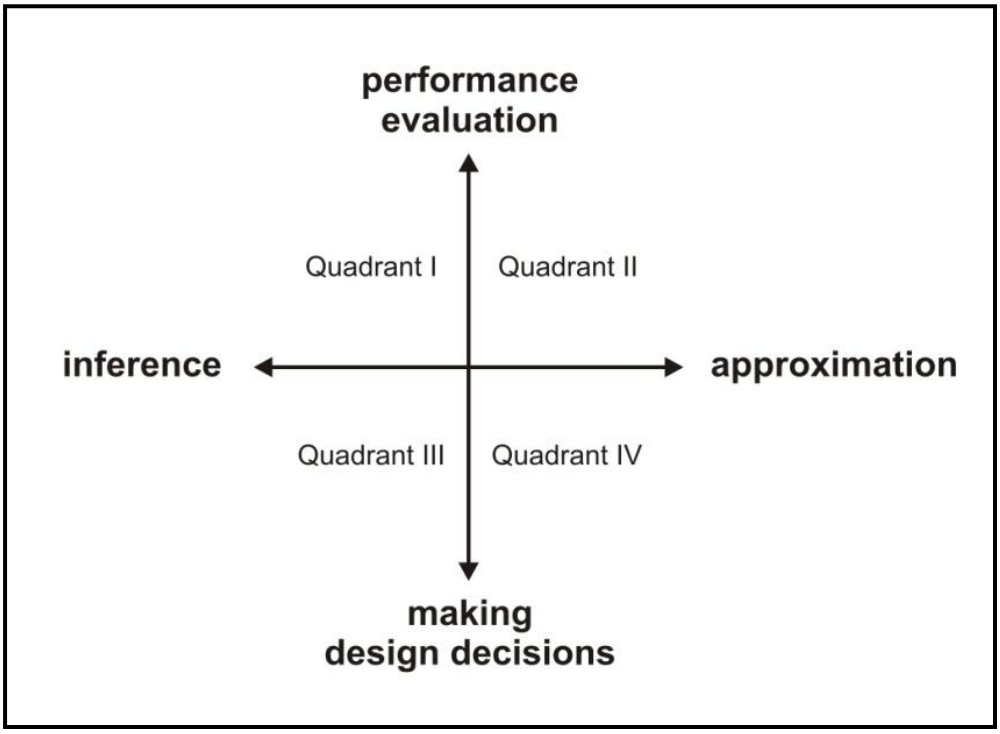


Figure 2-2 Types of usability tests with regards to aim of the test and methodological approach

Regardless of the nature of research and the motivations behind (see Figure 2-2) representative sampling and heterogeneity of user population are issues to be keen on for obtaining plausible results, unless the only function of observations is to inspire usability experts who rely heavily on their expertise for anticipating usability flaws. However, it should be noted that when a valid inference is to be made with the results of a usability study, control over factors pertaining to sampling that may affect test results becomes even more vital.

Although the main discussions in sampling literature concentrate on the discussions on sufficient sample size to discover the majority of usability problems (see Caulton, 2001 for a review), the probability of experiencing usability problems in a user test seems to be related with experiential factors. Therefore, all types of homogeneity assumptions, regarding age, gender, occupation, experience may prove to be inaccurate. If this is the case, then, even diversity and significance of the problems observed in a discount situation may not be plausible unless the sample is checked for serious biases in terms of expertise levels of the participants involved. With a small sample size even some of the most serious problems may not be encountered by the participants if the sample is heavily skewed in terms of experiential factors.

In the following section the problem of representative sampling in usability research will be discussed.

2.6. Problem of representative sampling in usability research

Usability studies that are characterized by user involvement are mostly non-experimental, that is, observational in nature (Nielsen, 1993), and are carried out for formative or summative purposes. Generally speaking, the primary aim is to diagnose usability problems in the former and to 'measure' performance in the latter. Regardless of the nature of research and the motivations behind, representative sampling is an issue to be keen on for obtaining plausible results, unless the only function of observations is to inspire usability experts who rely heavily on their expertise for anticipating usability flaws. For summative studies, representative sampling is even more vital since observations are supposed to lead to absolute statements about the usability of the system being investigated.

Although, the need for representative sampling finds support in literature, suggestions about factors to be considered are divergent. Furthermore, methods and techniques for obtaining a representative sample are not concretely put.

Nielsen states that “sample should be as representative as possible of the intended users of the system” (1993, p. 175). In order to achieve this, for the systems with large intended populations, anyone can be a participant; but age should be considered if old users are targeted and gender was found out to be significant in some cases. He further adds that novice – expert dichotomy was useful as a main distinction based on experience and in many cases both groups should be involved. He establishes the dimensions of user experience as computer experience, experience with the particular system, and domain knowledge. Finally, he adds that some “less immediately obvious” factors such as basic abilities were known to play a role. Chapanis lists the “human characteristics that are important” (1991, p. 375) as sensory capacities, motor abilities, intellectual capacities, learned cognitive skills, experience, personality, attitudes and motivation. Dumas and Redish (1993) suggest that “[d]eveloping a good profile of users should be a joint effort of the marketing department, usability specialists, and product designers” (p. 120) and if, for example, a system’s target is “mid-to large-size corporations…we will want to look for people who work in mid-to large-size corporations” (p. 121). They further add that experience and motivation are two important factors to explain differences among people, and propose a similar construct of experience with Nielsen (1993). The experiential factors to be considered are listed as: *work experience, general computer experience, specific computer experience, experience with the particular product, and experience with similar products* (p. 122).

Some of the approaches that are common in the studies reviewed above may be challenged in order to arrive at an alternative way of looking at the issue of representative sampling.

2.7. Alternative approach to the issue of representative sampling

First of all, a common attitude is exhibited in the sense that how experience is considered as an important factor and how it is defined. Experience is usually, if not always, defined as quantity, frequency and duration of participation to a task, interaction with a class of applications, a specific application, or computers in general. Such a construct is valuable and has practical appeal to present the multidimensionality of experiential differences. Moreover, such information is readily available and may be very helpful in discount situations. Nevertheless, it is better to treat such information to draw a coarse distinction between user groups. The problem of defining experience in such terms arises when experience is treated as a predictor of performance, as a confounding variable, or as a substitute for a variable representing the transformations occurred during learning process. Two users who have been using cellular phones for five years cannot be assumed to have the same level of expertise in using cellular phones. People certainly differ even after they attend a formal learning process; to the extent of knowledge and skills they acquired (Ackerman and Humphreys, 1990), which is actually one of the motives behind the study of individual differences. If such an approach to experience could be sufficiently valid, then no examinations would be necessary for monitoring people who attend educational programs.

Secondly, conventional approach to representative sampling does not overlap with the notion of *individual differences* in the way that is tried to be represented here. As far as the professional practice of usability research is considered, the measures of user performance do not satisfy the aims of the projects most of the time. Therefore, together with this basic area of interest, other aspects such as user satisfaction and usefulness are successfully integrated to concept of usability. With such an attitude, it is certainly good practice to have a sample of participants that matches the targeted consumer profile. However, if the research is focused especially on the objective measures of user performance, then representation of the consumer profile by a sampling scheme based on socioeconomics and demographics loses its vitality and plausibility.

A better conceptual position for identifying the attributes that directly influence performance should be looked for in order to ensure validity, even in commercial projects where the researcher is only interested in observing user performance. The concept of *expertise* rather than *experience* seems to be a proper starting-point for this purpose, given the fact that it emphasizes the acquisitions of individuals but not what is experienced. Expertise may briefly be defined as "aspects of skill and general (background) knowledge that has been acquired…" (Freudenthal, 2001, p. 23).

In the next chapter an approach based on expertise as defined here will tried to be constructed.

CHAPTER 3

3. GENERAL INTERACTION EXPERTISE

3.1. Definition of General Interaction Expertise

In a usability test, most of the time, if not always, participants experience a novel situation. In other words, either a new interface is being tested or participants are asked for completing novel tasks with a familiar interface. It is observed that participants try to grasp designer's model by navigating within interface and trying to complete the tasks assigned to them. Some participants may predict the model with quite ease before a thorough experience; while others may never form a working model of the system that conforms with the actual model and keep experiencing problems.

Therefore, in essence, in usability tests participants are asked to adapt to a novel interaction situation. As it is thoroughly discussed in Chapter 2, it is argued that a test participant's expertise level acquired by experiencing a diversity of interfaces is one of the most determining factors that affect how s/he copes with this novel situation. Term suggested for this construct is General Interaction Expertise (GIE) (Berkman & Erbuğ, 2005), and may be briefly defined as:

> **General Interaction Expertise (GIE)** is a general proficiency acquired by experiencing several interfaces, that helps users to cope with novel interaction situations.

3.2. Triadic model

In this study, the model suggested in Figure 3-1 will be utilized for comprehending the relationship between what is experienced (experience) and manifestations of what is retained (GIE)— i.e. expressions of permanent cognitive changes, as actual performance and self-efficacy belief.

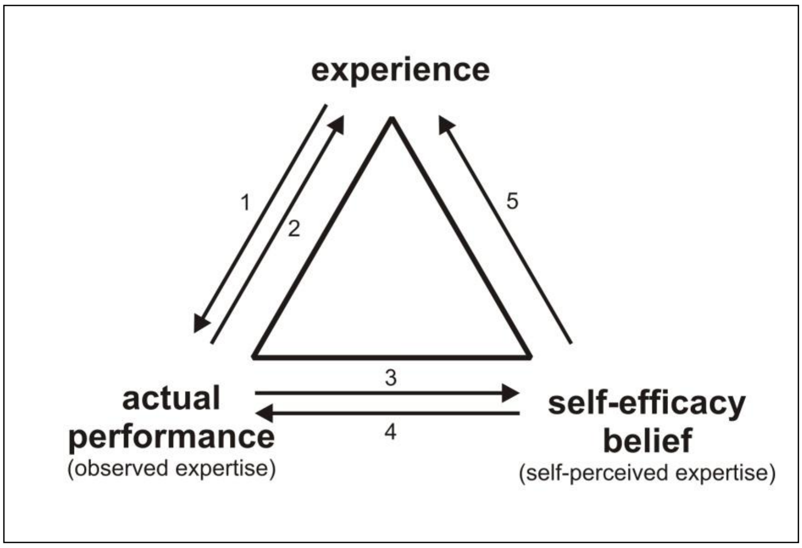


Figure 3-1 Triadic model of experience and components of expertise

This triadic model is in line with Bandura's social learning theory (1986). Before going into detailed discussion of the reciprocal relationships among the components of this model, the concept of self-efficacy should be briefly discussed.

The concept of 'self-efficacy' proposed by Bandura (1986) is frequently utilized to measure and even predict performance. According to Bandura, individuals possess a self system that enables them to influence their cognitive processes and actions. Therefore, "what people know, the skills they possess, or what they have previously accomplished are not always good predictors of subsequent attainments because the beliefs they hold about their capabilities powerfully influence the ways in which they will behave" (Pajares, 1997). In line with this view, researchers developed many scales that targeted 'computer self-efficacy' (e.g. Murphy, Coover and Owen, 1989; Compeau and Higgins, 1995; Quade, 2003; Barbeite and Weiss, 2004; Torkzadeh and VanDyke, 2001).

Suggested as 'more than just a mere reflection of performance', the concept of 'self-efficacy' was considered as a framework for defining the construct that will form the backbone of the scale under development.

3.3. Self-efficacy[2]

3.3.1. Definition

While discussing what is excluded and what is included to the term 'self-efficacy' Bandura asserts that self-efficacy is more than the possession of the required underlying skills for completing a particular task (1986). He maintains that "competent functioning requires both skills and self-beliefs of efficacy to use them effectively" (p.391). Therefore, self-efficacy is proposed as a generative entity that makes it possible to use skills, yielding a desired outcome, within various contexts. In this regard the concept is markedly different from outcome expectancies and can be delineated as an individual's self-belief in attaining a certain level of performance. However, Bandura views self-efficacy as a functional mechanism rather than just a self reflection on one's own capabilities.

> Self-percepts of efficacy are not simply inert estimates of future action. People's beliefs about their operative capabilities function as one set of

[2] This section is mostly based on Bandura's seminal work *Social Foundations of Thought and Action: A Social Cognitive Theory* (1986), where he situates the concept of self-efficacy within a broader framework.

> proximal determinants of how they behave their thought patterns, and the emotional reactions they experience in taxing situations. Self-beliefs thus contribute to the quality of psychosocial functioning in diverse ways.
>
> (1986, p. 395)

Stemming from this argument, it is suggested that self-efficacy partly determines which actions are undertaken and which social milieus are involved with. Therefore, as self-efficacy about a domain starts to grow, through its effects on choice behavior, it starts to determine what is experienced and what is avoided by the individual, partly influencing the course of personal development. It may be suggested that as self-efficacy beliefs are strengthened individuals may feel more motivated to get involved with the corresponding activities.

Another effect of self-efficacy beliefs is about breakdown conditions. It is argued that individuals with high self-efficacy beliefs do not give up easily when faced with obstacles and may even expend greater effort as they may tackle the problem as a challenge. Thus, it is asserted that individuals with strong self-efficacy beliefs tend to invest more effort and persist more in sustaining it.

A third effect of having strong self-efficacy beliefs is on the efficiency in converging cognitive resources on accomplishing the task at hand. Individuals with low self-efficacy tend to concentrate more on their limitations and shortcomings when they cannot proceed. Strong self-believers, on the other hand, concentrate on how to solve the problem and put more effort in dealing with ‘external’ problems. Furthermore, it is argued that high self-efficacy is related with causal thinking.

As a result, setting it aside from individuals 'actual capabilities', self-efficacy is a self-influencing mechanism, affects what actions people engage with, how they behave and how they act under stress or in situations of breakdown.

Proceeding from this general conception of self-efficacy and related mechanisms that stem from Bandura's cognitive theory, it may be proposed that a user with strong self-efficacy regarding interaction may be expected to have a tendency to use digital interfaces more often.

3.3.2. Sources of self-efficacy

Dwelling on the sources of self-efficacy perceptions are crucial for the definition of a construct that embraces the acquisition process, thus linking the self-efficacy based construct with the previous definition of General Interaction Expertise.

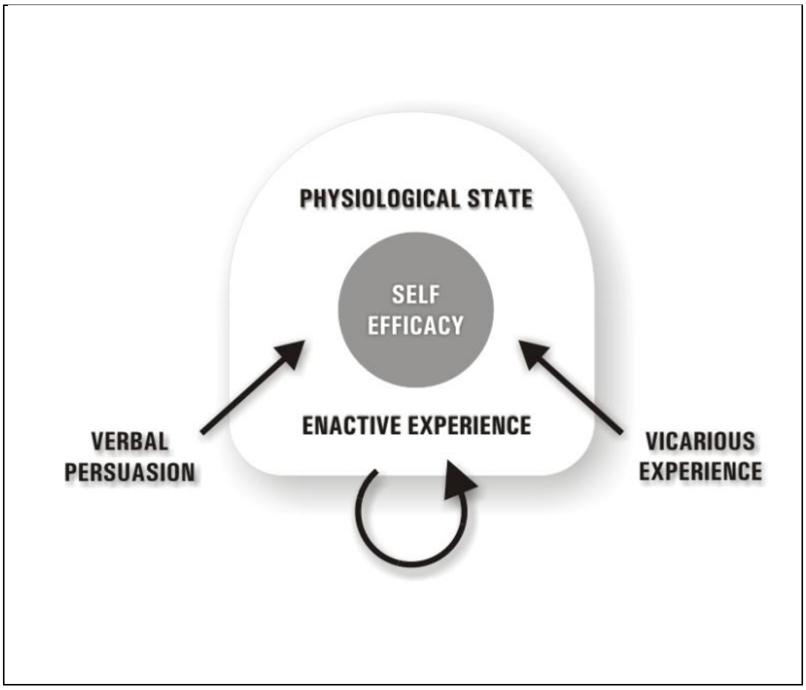


Figure 3-2 Internal and external sources of self-efficacy

The primary source for any self-efficacy belief is the enactive experience, where the individual experiences the domain. Bandura (1986) calls such experiences 'authentic mastery experiences'. Episodes that lead to success are deemed to strengthen the self-efficacy beliefs and poor experiences lower them. Furthermore, Bandura suggests that repetitive experiences that alter self-efficacy perceptions are slightly affected by rarely occurring negative outcomes. Therefore, as self-efficacy reaches to a certain level it becomes immune to disproving evidence. Together with this gain of robustness, beliefs tend to be generalized to other domains that are similar in character. Therefore, during the acquisition of GIE, experiences with products not only result in strengthening of a specific self-efficacy belief but also lead to construction of a generalizable form of self-efficacy. Marakas, Yi and Johnson (1998) discuss this issue in the case of computer self-efficacy and suggest that several application specific computer self-efficacy beliefs (A/S) form the General Computer Self-Efficacy[3].

Another source of self-efficacy is vicarious experience. Individuals may also base self-efficacy beliefs on other individuals' successful experiences. Furthermore, in cases where there are no absolute measures of success and failure vicarious experience serves as follows:

> When factual evidence for performance adequacy is lacking, personal efficacy must be gauged in terms of the performances of others. Because most performances are evaluated in terms of social criteria, social comparative information figures prominently in self-efficacy appraisals.
>
> (Bandura, 1986, p. 399)

According to Bandura, verbal persuasion is another method to alter or destroy an individual's self-efficacy belief. It is argued that it is harder to alter than to undermine an individual's belief permanently by verbal persuasion. Together with vicarious experience, this source frames the social facets of self-efficacy.

The last source is termed as physiological state and is related with self-monitoring of somatic responses in taxing situations.

---

[3] This conception of the acquisition of General Computer Self-Efficacy is again in line with the point mentioned in footnote 3. This similartiy in structuring the acquisiton process makes it easier to contain the self-efficacy concept.

> Because high arousal usually debilitates performance, people are more inclined to expect success when they are not beset by aversive arousal than if they are tense and viscerally agitated. Fear reactions generate further fear through anticipatory self-arousal.
>
> (Bandura, 1986, p.401)

This source of influence may be utilized to establish the interrelations of the concept with anxiety-related constructs.

Although Bandura does not offer such a dichotomy, these 4 sources may be formulated as internal and external (social) sources of self-efficacy.

Proceeding from this general conception of self-efficacy and related mechanisms that stem from Bandura's cognitive theory, it may be proposed that a user with strong self-efficacy regarding interaction may be expected to have a personal history of interaction where positive experiences are dominant, tendency to use and learn new digital interfaces more often, exhibit persistent behavior in breakdown situations, and not to exhibit self-blaming behavior in case of an error.

3.4. Construction of GIE

In order to discuss how GIE is constructed, each link between the elements of the triadic model should be examined.

3.4.1. Experience - Actual performance (1)

The suggested relationship between experience and actual performance (see arrow 1 in Figure 3-4) is tried to be illustrated by exploiting the elaborated taxonomy suggested by Smith (1997).

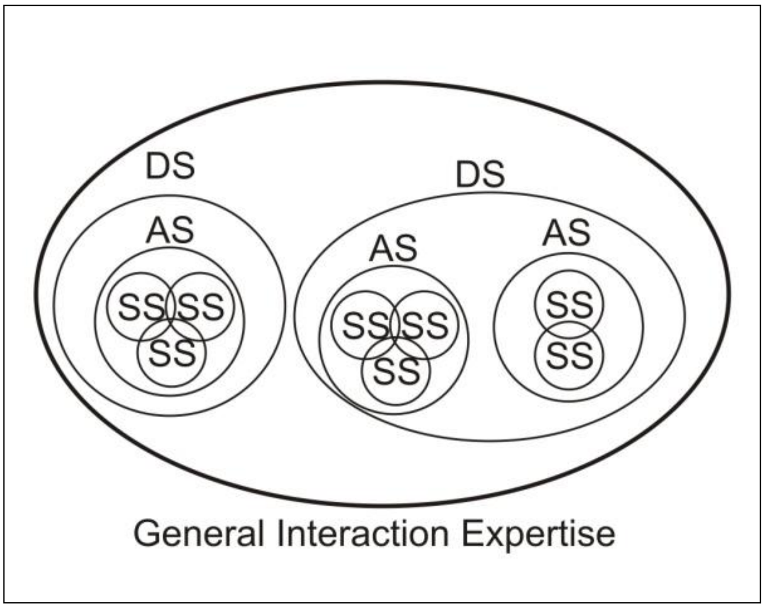


Figure 3-3 GIE, domain specific knowledge, application-specific component and system-specific component

It may be suggested that as individuals interact with a specific product they acquire a system-specific component of expertise (SS). After experiencing a number of similar systems for carrying out the same task—i.e. listening to music—an application-specific component (AS) of expertise is formed. Therefore, as people use specific systems with similar functionalities they acquire an AS together with individual SS components. Domain-specific knowledge (DS), on the other hand, consists of all the knowledge and skills required for carrying out a specific task. For example, etiquette of unmediated face-to-face communication may be situated within DS of communication.

Coming across a variety of SS, AS, and DS, several schema-based expertise (see Preece, 1994) are acquired, which help individuals to manage known and novel but familiar systems. Even if users face a totally novel application area, their expertise help them to orientate to the new system, provided that prior expertise acquired bear sufficient commonalities with the novel situation.

Therefore, although it was illustrated as if separate areas of AS and DS do not overlap in Figure 3, they actually do in reality. Moreover, the areas of intersection among separate areas of SS are larger than depicted.

This taxonomy is further clarified with a concrete example about using a washing machine in provided in Table 3-1.

.

Table 3-1 Using a washing machine with a digital interface

GIE	**Interaction**	Power on/off pictogram, navigating through menu structure, how cancel button functions...
DS	**Washing garments**	Procedure of washing, effects of temperature on textile and dyes, how to spare hot water, how to identify a well-washed cloth…
AS	**Washing with a machine**	Certain controls and displays specific to washing machines, functional model of washing machines, how to save energy, safety precautions …
SS	**Washing with a specific model of washing machine**	Program A, Program B, specific pictograms, menu hierarchies, procedures, key combinations …

3.4.2. Actual performance – experience (2)

The relationship between experience and expertise is suggested to be reciprocal one (see arrow 2 in Figure 3-4).

It may be argued that as an individual's expertise observed to be improved over time, a social image will be formed and probability of coming across with novel interaction situations may eventually increase. For example, if an individual is known to be good at handling novel interaction situations, individuals may start to consult her/him frequently. Thus, if an individual's observed expertise becomes prominent it may affect what will be experienced by her/him. On the other hand, if an individual is observed to be a poor performer then other individuals will not ask for help or encourage the individual to get involved in novel interaction situations.

3.4.3. Actual performance – self-efficacy (3)

As mentioned earlier, as individuals experience a diversity of interfaces they form a self-efficacy belief (see arrow 3 in Figure 3-4). This belief may be strong or weak depending on how the outcome of the experience was perceived by the individual. In other words, an individual's performance in novel interaction situations will be reflected in the form of self-efficacy belief.

3.4.4. Self-efficacy – actual performance (4)

As individuals grow self-efficacy beliefs about interaction, their actual performance with interfaces are influenced through several mechanisms (see arrow 4 in Figure 3-4). As discussed earlier, people with a strong self-efficacy belief are good at overcoming breakdown situations and converging cognitive resources to problem solving. People with low self-efficacy may tend to get frustrated easier, ask for help or may be prone to quit when confronted with a problem.

3.4.5. Self-efficacy – experience (5)

Individuals with strong self-efficacy beliefs with regards to interaction are expected to extensively learn and use new digital interfaces and to frequently get involved in challenging interaction situations. Individuals with a low self-efficacy may choose not to use digital interfaces and try to avoid challenging interaction situations as much as possible.

3.5. Actual performance and self-efficacy as manifestations of GIE

As defined by Cronbach and Meehl (1955), a construct is an attribute postulated to be possessed by individuals and reflected in behavior. It is developed “generally to organize knowledge and direct research in an attempt to describe or explain some aspect of nature” in a scientific inquiry (Peter, 1981, p. 134). It is only possible to make inferences about the attribute by examining its surface manifestations. Therefore, constructs can be observed indirectly.

As depicted in Figure 3-4, GIE was treated as a construct, which is manifested in actual performance and self-efficacy beliefs. Although it was mentioned that there is a reciprocal relationship between experience and expertise (see Figure 3-4, treating experience as a manifestation of GIE is methodologically inappropriate since ‘what is experienced’ is not a reflection but one of the causes of GIE in the first place.

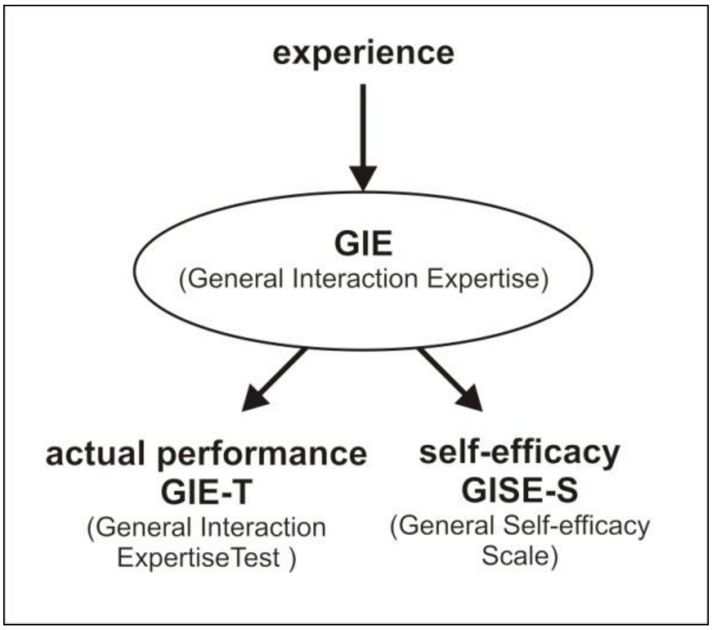


Figure 3-4 The construct of GIE and its main cause and its manifestations.

3.6. Measurement of GIE

According to the results of a brief literature review it was found that there are 4 main measurement approaches for studying constructs that target some sort of expertise related with the use of technology.

3.6.1. Actual tasks

In this approach, respondents are asked to perform certain tasks under controlled conditions. Although, it resembles the style of measurement adopted in apparatus tests the aim is usually to test the subject's proficiency of a particular software package.

It is not a widely resorted technique (e.g. Bunz, Curry and Voon, 2006; Kay, 1993). Unlike the apparatus tests suggested in Chapter 4, whether subjects can complete certain everyday tasks with an actual software package is observed. Thus, the aim is not to have a standardized test to gauge users' expertise in various research conditions but to utilize results mostly for personnel selection. In the literature, measuring expertise with actual tasks in order to explore its effect on other factors is not a frequently witnessed approach.

3.6.2. Verbal tasks

In the employment of verbal tasks respondents are asked to answer certain questions that aim to test computer related knowledge. Items of such tools mostly resemble written examinations or multiple-choice tests. Such tools are mostly applied in educational settings for measuring achievement (e.g. Jones and Pearson, 1996; Cassel and Cassel, 1984) of students.

Most of such tests are not standardized and applied in an *adhoc* manner by teachers in the form of classroom examinations. However, there are tools composed of standardized verbal tasks (see Cassel and Cassel, 1984).

3.6.3. Frequency and diversity of experience

When the effect of experience related with technology use on another phenomenon is explored, questions that target frequency and diversity of experience are widely utilized. Respondents are asked to report frequency and opportunity to use computers, diversity of computer experience (e.g. Bunz, 2004; Kinzie, Delcourt and Powers, 1994; Igbaria, et al. 2001) or similar technologies.

As it was discussed, although this approach looks very straightforward it is quite problematical. Such tools often neglect that frequency and diversity of experience is a necessary but not sufficient condition for a high level of computer literacy. For this, it is not a proper way of studying acquisition. Despite its methodological problems, the fact that such data may easily be gathered seems to appeal researchers.

3.6.4. Attitudes

Measures based on self-perception are often utilized in order to have an idea about theoretically impossible to observe traits. Respondents are asked to report their self-perceptions of related constructs (e.g. Loyd and Loyd, 1985; Murphy, Coover and Owen, 1989; Compeau and Higgins, 1995). By concentrating on attitudes researchers may gather information that may not be observed or measured without the collaboration of individuals.

Within these possibilities, given the research model adopted in this study, which is based on social learning theory, a scheme that consists of *actual tasks* and *attitudes* is suggested. Furthermore, such a scheme is in line with the aims of the study, and it is possible to form a triangulation by adopting two different approaches in measurement.

Although tests that include verbal tasks were considered during the development of the paper-based component, as an alternative to apparatus tests, inherent problems related with verbal tasks rendered them inappropriate. These problems were discussed in Chapter 4.

Besides the theoretical concerns, a measurement scheme consisted of one observational tool and a paper-based component had some practical consequences with regards to the employment of tools in real-life settings as well. These will be discussed in Chapter 6.

In Chapter 4 and 5 theoretical backgrounds, development processes and reliability/validity studies done for both tools were discussed in detail.

3.7. Potentials of measuring GIE

Below, the branches and types of research that would benefit from this method are suggested. For each branch, fictitious research designs were provided to exemplify a variety of possible uses of the tool.

3.7.1. For basic research

If GIE levels of participants would be determined with sufficient accuracy, it may open up the possibility to conduct research on various fields where expertise levels of participants should be controlled or manipulated.

Examples:

- An observational study that investigates how users behave in certain breakdown situations will be conducted. The tool may be utilized to check whether sample population is approximately normally distributed with respect to GIE since researchers believe that experience plays an important role in error handling.
- An experimental study is going to be conducted to discover the effects of expertise level on recognition and comprehension rate of iconographic and alphanumeric feedbacks. Here a 2 x 2 factorial design may be employed and the tool may be used to divide the sample into four:

Table 3-2 Allocation of participants

	High GIE group (N/2)	**Low GIE group (N/2)**
Iconographic feedbacks	N/4	N/4
Alphanumeric feedbacks	N/4	N/4

In an explorative study, how people discriminate between ‘user-friendly look’ and ‘childishness’ is investigated. Levels of GIE, together with many other attributes that are likely to be in charge, may be explored in accordance with participants’ perception of visual styles.

3.7.2. For applied research

Examples:

A totally novel mode of interaction, based on converting hand and body gestures to commands, is being researched. Although it is believed that this is a more natural way of control, researchers would like to find out whether this interaction type could be applied to familiar products without sacrificing efficiency. In order to explore the effects of 'negative transfer', the tool may be used to select participants with a considerable amount of expertise in conventional modes of interaction, thus more likely to experience negative transfer.

A research is conducted for exploring the maximum number of visual feedbacks that could be communicated to users concurrently, without causing information overload. Researchers would like to show that this limitation is determined mostly by the capacity of working memory rather than experience with interfaces.

3.7.3. For design research[4]

In applied situations where the aim is to guide the design process of an interface, the tool may be used to select appropriate participants.

[4] It seems impossible for a single measurement tool to answer the needs of every type of research. Therefore, it is feasible first to generate an eloborate tool suitable to basic and applied research. Consequently, a simplified version may be derived by comprimising methodological strictness to an extent, to arrive at a technique that will be easily applied in discount situations where resources are not in abundancy.

Examples:

In a design project, at certain phases of the process user tests are required to make sure that successive design decisions do not hinder usability of the product. In a longitudinal study of this sort, the tool may be utilized to guarantee that sample populations do not differ much in respect to experience with interfaces.

A focus group is planned for gathering comments and suggestions for a new interface. For a pool of creative ideas to be formed, research team is specifically interested in opinions of 'unbiased' users who do not have much experience with conventional interfaces

3.7.4. For projects done under contract

In projects done under contract, the tool may be used as a means of verifying assumptions about sample.

Examples:

A firm recently working on a new microwave plans to promote this model by emphasizing its ease of use. They would like to check whether the prototype can be effectively used by everyone. In this study the tool may be used to identify people with quite low GIE and include them to the sample population.

A home electronics firm is planning to compare one of their products with another product on the market. They would like to find out whether their design is more usable or not. In this case a two-sample research design may be applied. Ensuring that participants in both groups are almost equally-distributed with regards to GIE would be helpful in eliminating the effect of expertise in observed performances.

CHAPTER 4

4. MEASUREMENT OF ACTUAL PERFORMANCE

In this chapter two apparatus tests that are developed for identifying expert behavior by analyzing the actual performance of individuals in standardized interaction situations are discussed. Before presenting details about the development process of the apparatus tests a theoretical foundation is provided based on automatic – controlled processing dichotomy, which will be discussed. Finally, results regarding both reliability and predictive validity of the tests were reported.

4.1. Automated processing

Everyday activities that people carry out are usually composed of automated processes. It is possible to handle such tasks while attending to another one. Such a process of automation is observed in many of the sensory-motor tasks that are practiced frequently. After a sufficient period of experience, even demanding cognitive processes are observed to become automatic (Preece, 1994). From information processing perspective the phenomenon may be explained with the theory of automatic and controlled processing. Automatic processes demand little effort, may be unavailable to consciousness, and maybe identified by their fluency; whereas controlled processes, tap a considerable amount of cognitive resources and are slower than automatic processes (Sternberg, 1999). According to Ackerman (1987), after sufficient practice under consistent task conditions, controlled tasks may become automatic. For consistent tasks, improvements in performance are limited with individual's sensory-motor capacity or motivation to perform better.

Even it has sprouted from a different school of thought; Activity Theory provides a similar explanation to the process of learning. According to Vygotsky (1978) when people get involved in an activity, they make plans that help them to formulate actions, which are meant to satisfy certain sub-goals. Actions, then, are actualized by a set of operations. After individuals gain certain expertise, actions and even whole activities are carried out as routine operations. However, when conditions vary, a simple operation will be handled as an Activity in itself (see Koschmann, Kuuti and Hickman, 1998 and Bodker, 1991 for a complete model).

Both theories have common points that give clues about ways of recognizing expert behavior:

- The extent of expertise gained by practicing a task may be predicted by whether the task is automated, still under conscious control, or both.
- After a certain level of automation is attained in a specific task, gains can be transferred to other tasks with similar conditions.

Therefore, sensory-motor fluency observed in an easy task with a familiar interface may be an observable indication of expertise. Individuals with a high level of GIE would have been gained expertise by practicing similar tasks and may be expected to switch to automatic behavior after a concise orientation period.

Based on theories discussed above, it is suggested that GIE may be manifested in two fundamental types of behavior, which are automatic loops of execution – evaluation (GIE_XEC) and controlled problem-solving (GIE_PS). In order to assess expertise by observing actual performance on tasks that target these two types of behavior, GIE-T that consists of two prototypic apparatus tests were developed.

4.1.1. GIE_XEC: Study I

The following set of heuristics guided the development process of GIE_XEC test:

- Task content should be neutral, so that prior knowledge specific to systems, applications and domains should not alter performance.
- Test should not contain tasks that require cognitively complex processes.
- Test should not be comprised of tasks that require novel modes of interaction.
- Test should be comprised of familiar sub-tasks in order to maximize the effects of experience with digital interfaces on performance.

An apparatus test was developed in accordance with the theoretical framework and criteria stated above. The task consisted of three simple sub-tasks, assumed to fall into execution and evaluation domains defined previously. Task content was deliberately reduced as to eliminate the direct effects of SS, AS, or DS. Task difficulty and novelty was tried to be adjusted to a level so that indications of automatic processing would provide a partial estimate of individuals' GIE for the specific case.

Test software

For the collection of keystroke latencies, a GUI developed with *Macromedia® Flash MX 2004* was utilized. The interaction was consisted of 3 virtual subtasks that required basic actions such as navigation among menu items, selection, and manipulation of fictitious variables. Software was able to log the following data.

Initiation latency (T_{INIT}) – time required for the system to load and initiate task screens in milliseconds.

Keystroke latency (T_K)– latency between last key release and present keystroke milliseconds.

Elapsed time (T_{NOW}) – time elapsed until corresponding keystroke ($T_{INIT} + T_{K1} + \ldots + T_{Kn}$) in milliseconds.

Keycode – codes for the key pressed (U: *UP*, D: *DOWN*, L: *LEFT*, R: *RIGHT*, S: *END*).

Users controlled the cursor with a standard key set of a laptop PC (see Figure 4-1). The buttons used and their functions were as follows:

Table 4-1 Keys and associated functions

Key	System response
UP	Cursor moves up unless restricted with a boundary
DOWN	Cursor moves down unless restricted with a boundary
LEFT	Cursor moves left unless restricted with a boundary/ Decreases a parameter
RIGHT	Cursor moves right unless restricted with a boundary/ Increases a parameter
END	Selects an item/ Confirms an action

Task was composed of 3 subtasks. In the first subtask, subjects were required to select the item *modify* (*değiştir*) within a 2x8 list (see Figure 4-1).

In the second subtask, subjects were required to select the red square labeled P by moving the cursor to the bottom right corner from an initial position of top left corner in a 4x4 matrix (see Figure 4-2).

Finally in the third subtask, 5 fictitious parameters were modified by increasing or decreasing the values until each of them are *50* (see Figure 4-3).

Figure 4-1 Task 1 – Main menu

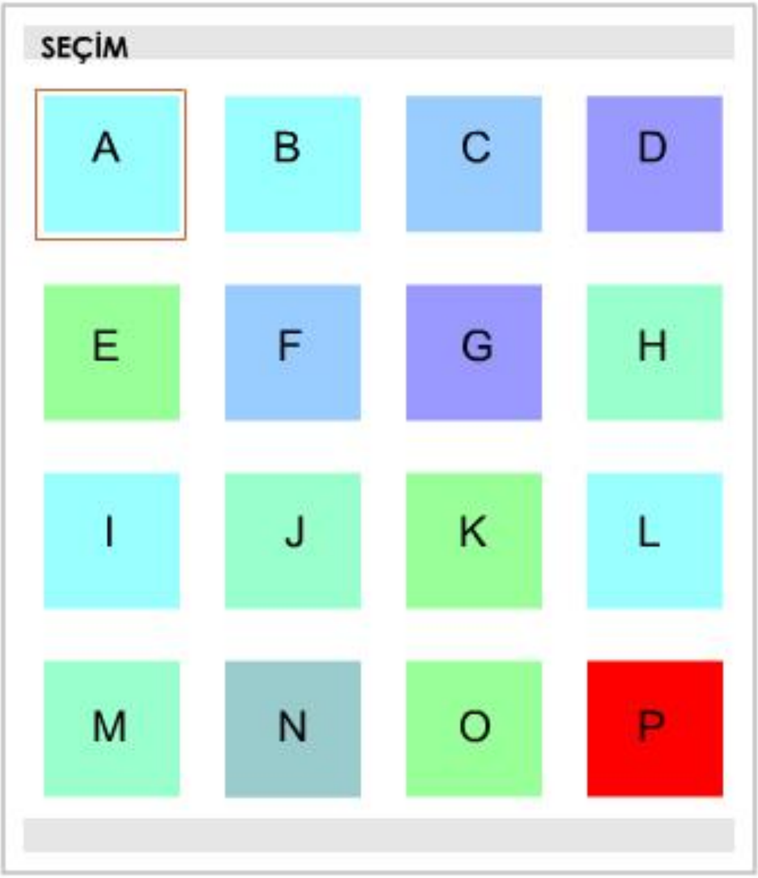


Figure 4-2 Task 2 – Choice

Subtask 1: Move the cursor to *modify* (*değiştir*) with *arrows* then select it by pressing *END*.

Subtask2: Move the cursor to square labeled P with *arrows* then select it by pressing *END*.

Subtask3: Increase/decrease each value with *LEFT/RIGHTt* then proceed to the next value by

Figure 4-3 Task 3 – Setting parameters

A laptop PC was used for the tests. Screen was checked for glare each time before a test session. Keyboard was positioned so that there was ample space for wrist support (see Figure 4-4). Keyboard settings *repetition latency* and *repetition speed* were set to minimum in order to avoid uncontrolled inputs with a single keystroke.

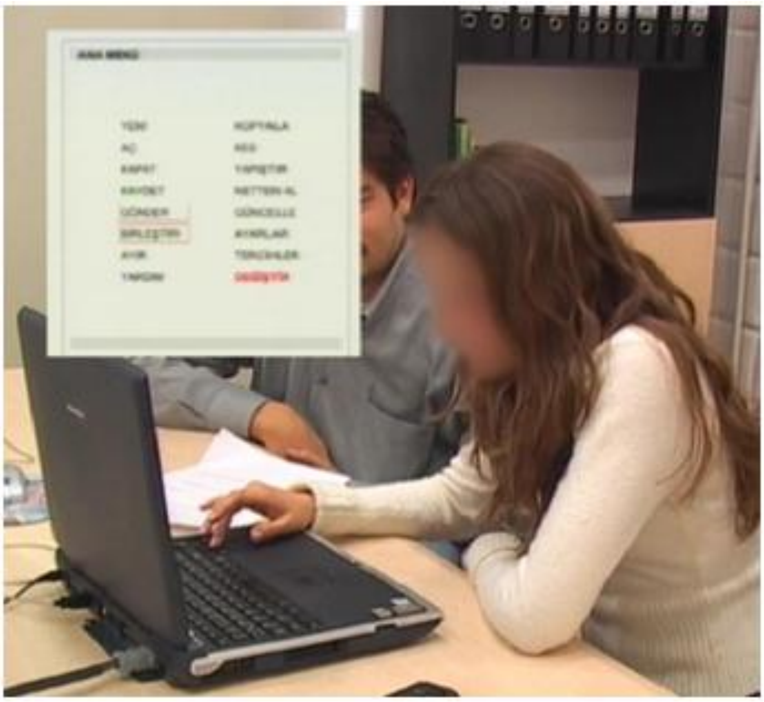


Figure 4-4 Test room configuration

Tests were conducted in a usability laboratory (METU – BILTIR) with a single observer. One portable digital camera fixed to a tripod, a scan converter, a digital V/A mixer, a boundary microphone, and a PC equipped with an encoder capable of recording real time mpeg files were used in recording.

Sample group consisted of 40 undergraduates studying in METU Department of Industrial Design (see Table 4-2). Quota criteria employed for sampling were gender and grade (see Table 4-2).

Table 4-2 Sample population

Grade	Gender	N
First	Female 5, Male 5	10
Second	Female 5, Male 5	10
Third	Female 5, Male 5	10
Fourth	Female 5, Male 5	10

$\sum N =$ 40

Subjects did not receive any extra credit for their participation. Recruitment was done by announcement and volunteers were drafted as subjects[5]. With this sampling profile, it may be argued that sample group was quite homogenous regarding age and educational level. Moreover, must courses on computer literacy are assumed to provide a basic level of computer skill.

Pre-test phase

- Before the tests, subjects were shown the observer room and the scene that would be recorded.

[5] The fact that subjects did not receive any extra credit may introduce non-respondent bias and volunteers were not representative of the whole population. However, if hypotheses are reviewed it is obvious that this even makes it harder to reject null hypothesis associated with H_1 to the extent that sample group may be assumed to be positively biased regarding computer literacy.

- Subjects were taken to the test room and informed about the camera that is shooting the scene.
- A brief description about the aim of the study was given without giving clues about what was expected or comments that might bias the subjects prior to test.
- Subjects were given exclusive instructions about the tasks, the functions of the keys, and procedures that should be followed in order to complete each task. Subjects were not told to follow a specific navigation pattern during subtask 1 and subtask 2.
- Subjects were told that the aim was to observe the natural behavior so that they should not pause for asking questions until a trial was finished and to avoid unnecessary actions.
- Subjects were told that none of their actions would be interpreted as right or wrong but interaction would be examined regarding the nature and style.
- Personal information such as surname-name, gender, year of birth, years passed in the university, and department was gathered.

Test phase

- Subjects were accompanied by an observer whom sat next to them.
- During performances all attempts of conversation was tried to be avoided.
- Each session was consisted of 6 trials of subtasks 1,2, and 3
- Before each trial, subjects pressed a key to confirm that they were ready to proceed.

- After each trial a non-task screen was displayed providing information about trial number.
- After the last trial subjects were prompted that the test was over.

Post-test

After the tests were done log files were converted for further analyses and video files were analyzed for gathering *orientation* and *visual feedback* data. The following variables for each subject were utilized in the analyses.

Table 4-3 Variables gathered

Variable	**Gathering method**	**Data type**
Gender	Pre-test questionnaire	-
Year of birth	Pre-test questionnaire	-
Orientation	Video analysis	Ordinal variable[6]. How subjects orient their hands most of the time on the keyboard. 1: single 2: double 3: triple 4: double hand
Visual feedback	Video analysis	Discrete scale variable. How many times subjects get a visual feedback in order to locate a key.

[6] TNumbers assigned are not arbitrary. Ranking was done assuming that 1 is inferior to 2, 2 to 3, and 3 to 4.

Table 4.3 cont'd

Initiation latency	Automatic logging	Continuous scale variable in ms
Keystroke latency	Automatic logging	Continuous scale variable in ms
Elapsed time	Automatic logging	Continuous scale variable in ms
Keycode	Automatic logging	D,U,L,R,S Errors are logged between two *X*s.

Keystrokes were sorted in to 4 types of latencies. L_0 (Latency 0) was assigned to the first keystrokes in each subtask. Keeping with the Keystroke-level model terminology (Card, Moran, & Newell, 1980) this type of latency may be said to be consisted of the following latencies.

$$T_{L0} = T_{acquisition} + T_{feedback} + T_{homing} + T_{Key}$$

$$T_{L1,2,3} = T_{feedback} + T_{mental} + T_{Key}$$

L_1 was assigned to successive keystrokes with the same key.

L_2 was assigned to keystrokes after a transition from one key to another.

L_3 was assigned to keystrokes on *END*.

Following example illustrates how the grouping was done.

[screen is loaded] L, L, L, L, L, L, D, R, R, R, R, R, R, D, S [end of subtask]

Latencies for each group of keystrokes are L_0, L_1, L_2, L_1, L_2, and L_3 respectively. After obtaining the log files, all the keystroke data were grouped for each subject and each task data was checked with single axis scatter plots for outliers. Outliers were conservatively omitted in a manual fashion[7].

Table 4-4 summarizes the expected number of latencies for each trial.

Table 4-4 Expected frequencies for latencies

Latency types	**L0**	**L1**	**L2**	**L3**
Expected f for each trial	3	57	11	3
expected f for 6 trials	18	342	66	18

[7] Keystroke latencies should not be viewed as reaction times. Since each keystroke latency have the possiblity to contain a mental component only extreme outliers were accepted as outcomes of distractions and were discarded manually, by doing a cross-check with video files. The reason why median of each group was not chosen for expressing central tendency is the fact that it is not suitable for further statistics.

Mean latencies for each subject, keystrokes omitted/included and elapsed time were gathered as quantitative data.

In addition to these, observable data such as orientation and visual feedback were regarded as potential predictors of GIE and were included in the evaluation.

Results and discussion

Readily-observable data, namely *orientation*, *visual feedback*, and *# of keystrokes* are provided below (see Table 4-5). For two of the subjects (N13, 18), number of instances of visual feedback could not be detected due to fact that subjects blocked the view by inappropriate postures.

Table 4-5 Orientation, number of visual feedbacks and number of keystrokes recorded

N	Orientation	Visual feedback	# of keystrokes
1	2	21	437
2	3	29	439
3	1	46	468
4	2	33	436
5	2	28	449

Table 4-5 cont'd			
6	3	6	446
7	1	25	440
8	3	12	446
9	2	35	430
10	2	19	435
11	1	86	436
12	3	24	442
13	1	?	450
14	2	20	437
15	2	20	445
16	1	24	451
17	1	32	433
18	3	?	439
19	2	36	441
20	3	20	431
21	2	32	443
22	3	16	433
23	1	71	445
24	1	67	438
25	2	19	450
26	1	24	441
27	3	17	437

Table 4-5 cont'd			
28	3	26	445
29	2	29	438
30	3	32	440
31	1	29	438
32	4	5	435
33	2	22	436
34	3	20	433
35	1	27	433
36	2	33	461
37	1	51	448
38	3	25	442
39	3	19	454
40	3	8	441
	1: single 2: double 3: triple 4: two-handed		

Further evaluation of the data provides that there is a significant correlation between the type of orientation and number of visual feedback needed. Pearson's coefficient (r) was -.622 at the 0.01 level (one-tailed). This indicates a significant negative correlation between the variables, which is expected (see also Figure 4-5). For instance, while single fingered subjects require a vast number of feedbacks, two handed orientation (adopted only by N_{32}) requires much less. Therefore, both variables can be assumed as partial predictors of GIE on their own.

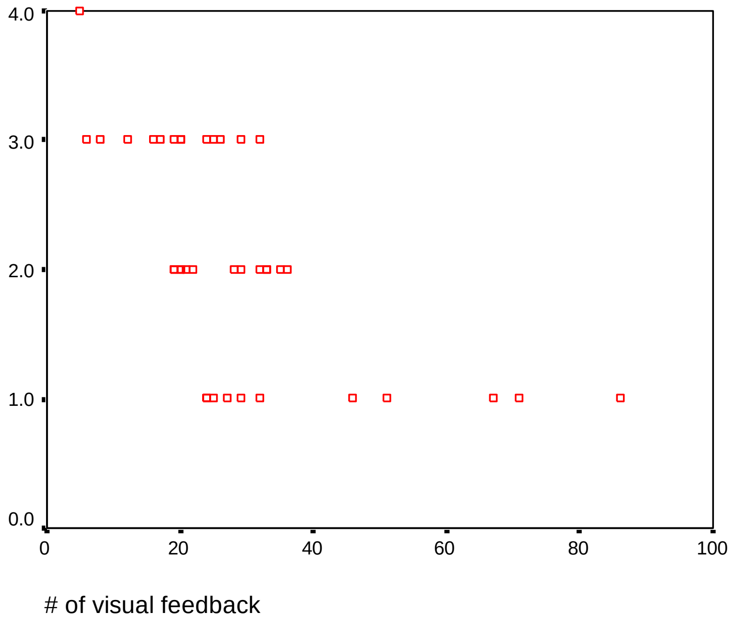


Figure 4-5 Scatter plot of orientation vs. #of visual feedback

To what extent readily-observable data and variables based on keystroke latency have a correlation is summarized in Table 4-6.

Table 4-6 Bivariate correlations (Pearson's r) of variables

	orientation	#of visual fb	L1	L2	L3	L0	SN
orientation	1.000	-.622**	-	-	-	-	-
#of visual fbs	-.622**	1.000	.140	.652**	.337*	.315	.299
L1	-.425**	.140	1.000	.404**	.352*	.292	***
L2	-.625**	.652**	.404**	1.000	.599**	.594**	***
L3	-.494**	.337*	.352*	.599**	1.000	.509**	***
L0	-.496**	.315	.292	.594**	.509**	1.000	***
SN	-.437**	.299	***	***	***	***	1.000

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).
*** Variables are not independent.

Two additional variables included were how subjects position their fingers on the controls (orientation), and number of instances of looking at the controls before a keystroke (# of visual fbs.). A further variable was calculated (SN) to represent the deviation scores regarding means for L0, L1, L2, and L3, since it was assumed in cases of automatic behavior, deviation should be minimal. However, it was concluded that high correlations among variables may render calculating SN unnecessary, since basic variables were likely to yield similar results.

4.1.2. Study II: Predictive validity

After revising the apparatus for bugs and operational problems, it was administered in a real usability test to see whether there is a considerable correlation between usability performance and any of the basic variables explored in Study I. User performance data was gathered during a user test for a dishwasher with a digital interface. Effectiveness across the task scenarios applied to a sample of 15 participants was assigned as the variable that represents user performance.

Table 4-7 Raw scores and correlations between values observed for each variable and effectiveness.

subject	L0	L1	L2	L3	mean elapsed times	std. of elapsed times	#of erro rs	#of visual fbs	effectiveness
1	805,45	200,54	329,62	551,26	22041,80	3303,82	22	26	**80**
2	700,15	166,44	316,17	464,31	18076,20	2000,71	12	1	**80**
3	1780,01	262,54	656,91	749,05	36459,00	6184,08	4	57	**65**
4	1192,84	202,94	597,12	598,99	29143,00	4659,21	15	NA	**40**
5	1301,18	226,30	656,50	847,55	29994,60	4175,38	8	NA	**20**
6	1143,95	245,00	611,02	728,64	29295,20	1816,02	4	54	**65**
7	3756,14	385,93	1514,83	1338,34	74839,60	14759,43	59	153	**20**
8	3395,76	302,30	1031,47	921,80	64363,20	22311,79	4	101	**0**
9	997,20	187,14	438,41	640,24	24088,20	1972,83	14	28	**50**

Table4.7 cont'd									
10	1595,74	210,77	617,81	511,72	29125,40	3678,56	15	60	**40**
11	921,09	232,59	493,45	683,97	27311,20	1946,38	11	14	**25**
12	879,10	183,04	372,96	480,42	20605,60	2250,32	6	28	**50**
14	1413,38	236,66	597,63	1190,02	30930,40	1773,17	8	30	**50**
15	934,96	190,44	488,50	573,08	23992,60	544,35	24	63	**60**
16	1493,52	189,31	593,20	1207,13	26927,60	1436,27	0	NA	**60**
r	**-0,66**	**-0,59**	**-0,66**	**-0,39**	**-0,68**	**-0,68**	**-0,17**	**-0,60**	

Significant correlations ranged from -0.59 to -0.68. The highest correlation was observed with *mean elapsed times.* This high negative correlation indicates that subjects who completed tasks faster were more successful in completing the tasks in the usability test. Although the correlation was quite high in the initial state, this finding should not be overestimated. It may be interpreted as an indication of a common factor that influences both apparatus test performance and user performance.

According to the initial findings, it may be argued that, performance in this test may confidently be represented parsimoniously by means of observed elapsed times. Although a strong net of correlations among keystroke-level variables were discovered in Study I, analysis on the level of individual keystrokes seems to add nothing to the predictive power and may be left aside for the sake of simplicity.

4.1.3. GIE_PS: Second apparatus test: Theoretical foundations

In the beginning of this chapter, it was stated that the measurement of actual performance could be based on tests developed to fit automatic – controlled processing dichotomy. Here, in this section, a collection of models of interaction were thoroughly reviewed in order to focus on controlled processing to be covered with an additional apparatus test.

Norman's Action Cycle

According to Norman (1988), human action consists of two main components. In order our goals to be fulfilled we should be able to perceive and evaluate the current state of the world. This is followed by a set of actions for changing the world so that our goals are accomplished.

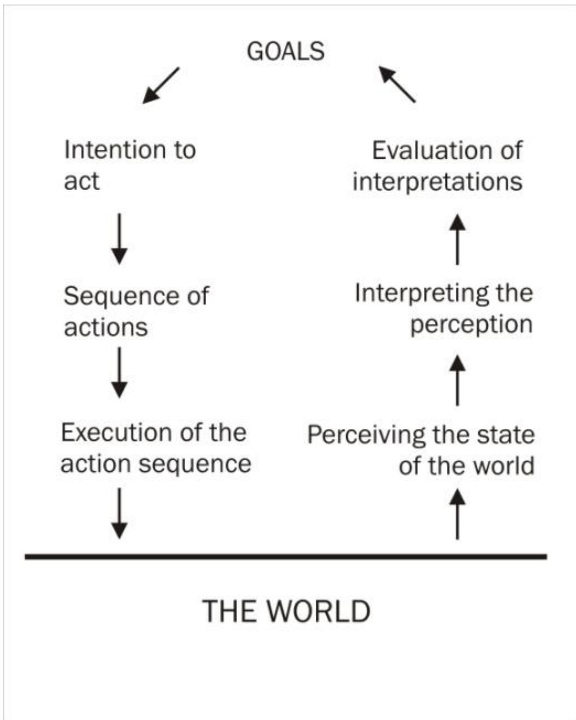


Figure 4-6 Task Action Cycle (Reprinted from Norman, 1998, p.47)

Therefore, the steps of the cycle presented in Figure 1 continuously follow each other until the "the world" is transformed so that our goals are satisfied. However, whether the flow is smooth or constantly interrupted, whether a single iteration is enough or the cycle is run many times depend on the characteristics of the components of interaction. On one end, cycle may be so internalized by the user that both concretizations of goals and interpretation of the world may be minimally crucial.

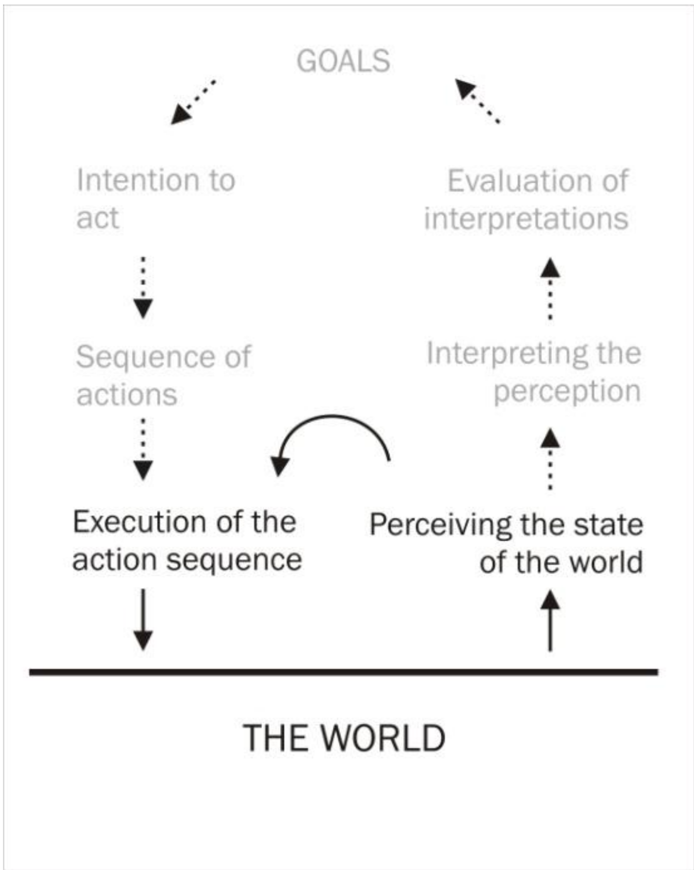


Figure 4-7 The Action Cycle by-passed

Taken to the extreme, executions may dominate the cycle, that is, automatic processing may take place minimizing even the need for perception in the form of feedbacks. In the first apparatus test (GIE_XEC), type of behavior tried to be addressed was fluency in such an automatic loop of execution – evaluation.

On the other extreme, there may be cases where *sequence of actions* may not be readily available, or “interpreting the perception” may not be possible. This usually occurs when people confront with serious problems with a known system, or when they came across with a totally novel interface. In such cases, translation of *intention to act* to a meaningful *sequence of actions* and to transform perceptions to evaluations may be problematic. With similar concerns, Sutcliff et al. (2000) propose certain elaborations which transform the model so that the level of detail is sufficient to discuss breakdown and learning situations.

In Figure 4-8, certain shortcuts and sub-cycles are suggested to embrace rather extreme cases mentioned above.

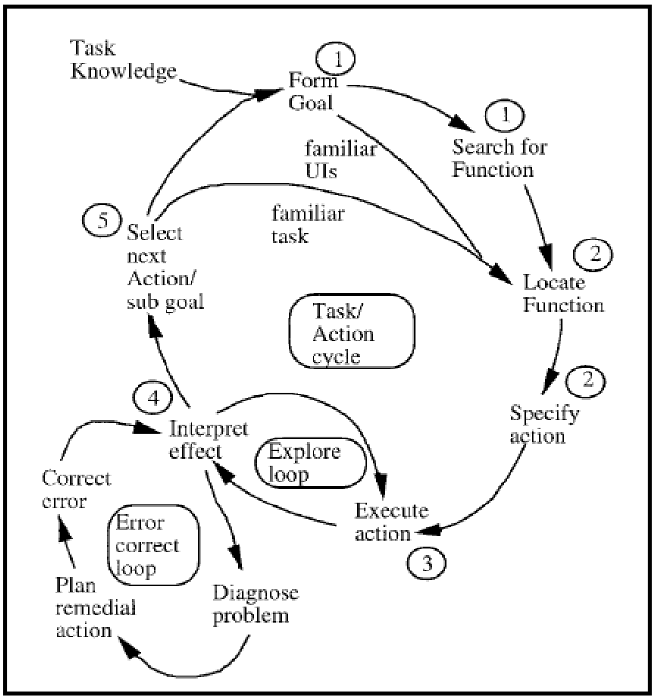


Figure 4-8 Task Action Cycle revised by Suttcliff et al. (2000, p. 45)

Problem-solving

Although they adopt a slightly different theoretical basis, Mack and Montaniz (1994) state that these extreme cases may be associated with quite different set of behaviors:

> A user experiences a problem when that user cannot accomplish some task because of the software tool being used, or can only do so with more difficulty than is expected or is acceptable. We assume a user has some goal (based on some task) to accomplish and that this overall goal can be broken

> down into a sequence of subgoals and actions appropriate for achieving each one. To the extent that these tasks are well-understood and practiced, we can characterize the goal-directed behavior as a routine cognitive skill. To the extent that the tasks or software interface are novel, we can characterize the goal-directed behavior in problem-solving terms and in terms of learning…
>
> (p. 301)

As opposed to "routine cognitive skills" commonly tapped in interaction with familiar systems, novel situations require problem-solving activity which at the end is terminated possibly with learning. As far as the elaboration suggested by Suttcliff et al. (2000) is concerned, this type of behavior is represented by *error correct loop* and *explore loop*. While discussing learning through experiences, Proctor and Dutta (1995) typify this problem solving – learning behavior with cases of learning to operate complex devices without instructions.

> Often, a person attempts to learn a device without the aid of instructions either because reading the instructions is perceived to be too time consuming or effortful or simply because the instructions accompanying the device has been lost.
>
> (p. 192)

It is evident that in a typical usability test this type of behavior is deliberately encouraged to see whether the product provides an intuitive mode of interaction. Therefore, it is possible to state that, in almost every usability test, participants are first confronted with a problem-solving activity, hopefully followed by a relatively smooth, uninterrupted *task-action cycle.*

Shrager and Klar (1986, ctd. in Proctor & Dutta, 1995) conducted an experiment to model the phases of learning where instructions are not available. After observing participants trying to cope with a quite novel interface, they defined the phases of the process as shown in Figure 4-9.

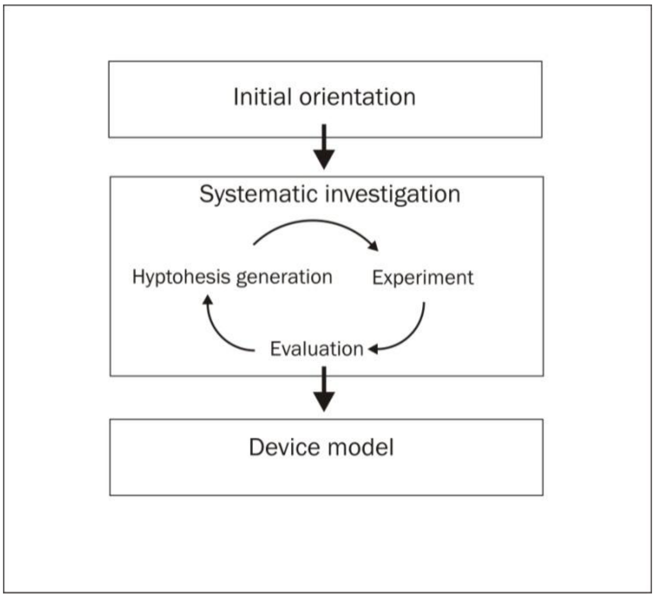


Figure 4-9 Learning without instructions (suggested after Shrager and Klar,1986)

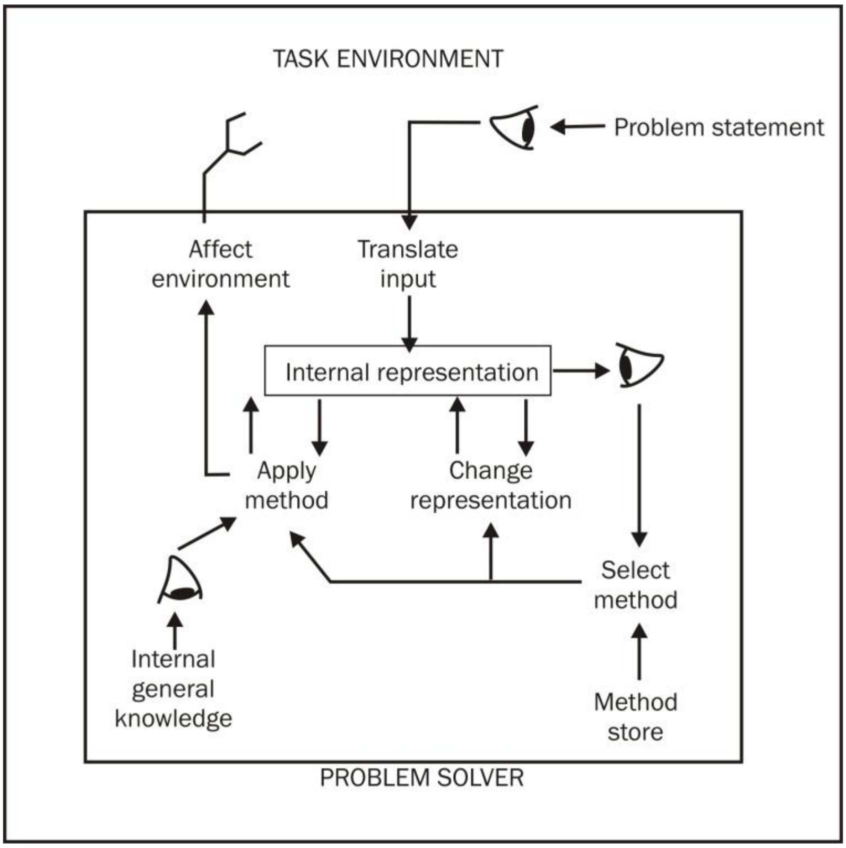
After an *initial orientation* phase where learn how to change device state, participants started to systematically investigate the system by generating hypotheses about ways of attaining task goals. These hypotheses were then tested and the ones that are verified helped participants to construct and refine the device model built so far. Therefore, in terms of Mack and Montaniz (1994), *systematic investigation* phase represents the problem-solving activity.

All the studies reviewed above mention some sort of problem-solving activity that takes place at some instances of interaction. This indicates that any research with an aim of exploring user expertise should essentially cover problem-solving type of behavior as an object of study.

None of the studies aim to study this phenomenon structurally by suggesting a cognitive model that underlies the process. However, in order to suggest 'what it takes to be an expert' in such types of behavior, firm links between observed actions and inner structures may be helpful. In this regard, the seminal work *Human Problem Solving* by Newell and Simon (1972) is worth an overview. Certainly, their definition of the term *problem* is totally in line with what is initially experienced by a participant in a usability test:

> A person is confronted with a problem when he wants something and does not know immediately what series of actions he [sic] can perform to get it.
>
> (p. 72)

After a problem is confronted the cognitive structure engaged with, is schematized in Figure 4-10.



Note. Eye indicates that input representation is not under control of inputting process.

Figure 4-10 General organization of problem solver (Reprinted from Newell and Simon, 1972)

According to the model, first problem solver translates the external problem definition into an internal representation. This representation forms the framework in which the problem solving will take place. In accordance with this representation a suitable method is selected. Application of the method, in turn, both affects the representation of the problem and the environment. At some instances the application of the method may be halted due to numerous reasons. In such cases, (1) a new method may be selected, (2) internal representation may be modified, or (3) the problem solver may give up.

Even though the suggested model may be criticized of presenting a reductionist perspective, it seems accurate in indicating the sub-mechanisms of problem solving, thus, providing clues about in what ways a user with a considerable expertise differ from a novice. Together with the apparent qualities pertaining to experts such as extensity and intensity of interface experience; efficacy in building internal representations when the problem is ill-defined and flexibility in exploring a diversity of methods to obtain the desired outcomes seems to be distinguishing qualities of expert problem solving. These two sub-mechanisms are unified under the term analytical skills by Lansdale and Ormerod (1994):

> Analytical skills are like the controlled processes […], in that they are highly flexible but require conscious thought before application. They allow user to understand how a task is performed with one interface, which may enable them to generalize their understanding to another interface and to modify aspects of their performance when the desired results are not obtained…
>
> (p. 164)

Furthermore, in line with Newell and Simon’s ideas, they state that both prior knowledge (*internal general knowledge* and *method store*) and ability to derive abstract knowledge (*translate input, select method* and *change representation*) out of that.

When it comes to everyday cases of problem-solving in interaction, another issue arises. Most of the time, the contents of the user's *method store* and the methods implemented within an interface may be different, or even conflicting. This is the same phenomenon described by Norman (1988) as *the gap between user's and designer's model*. It is assumed that as the user's experience with a diversity of interfaces deepens, the gap should become narrow and the overlap between two repertoires should be considerable. This is of course possible if one can speak of a unifying notion of interaction that is consistent enough, and is both available to designers and users. Therefore, one may expect that, as their experiences grow, users learn to successfully represent the arbitrary device models implemented within interfaces.

Development of the second apparatus test

As it was presented in Section previously the first apparatus test (GIE_XEC) consisted of a series of sub-tasks that aim to observe participants within a non-problem situation, where clear instructions were provided to eliminate problem-solving activity. The rationale behind the test was the assumption that as experience grows, familiar tasks are handled at the level of automatic processing, freeing valuable sources of higher cognitive facilities. Therefore, as a result of repeated exposure to similar familiar tasks of such as navigation, selection and modification; participants with high GIE would complete the tasks more fluently. Up to now, empirical findings seem to be in line with these major assumptions. Nevertheless, it is stated that performance at low level processing, on its own, would not be representative of the construct defined as GIE. Considering the theoretical background presented, a second test for the observation of problem-solving type of behavior seems necessary.

With such concerns, a second apparatus test (GIE_PS) was developed. The following criteria were considered during design in order test to measure what it intends to do:

- Goals states and current state of the device should be apparent to the participants. Participant's performance should not be hindered while trying to understand the goal state or compare it with the current state.

- Task should not require domain knowledge or a specific ability. Task to be completed should be neutral regarding other types of individual differences that are unrelated with GIE.

- Task should be easy to complete without the interface. If the task would be handled in an unmediated manner, all of the participants should be able to complete it (e.g. with paper and pencil, or verbally). The core of the problem should be related with grasping the device model implemented in the interface.

- The problem-solving activity should target relevant sub-mechanisms. The task difficulty should be related with how the problem is represented, flexibility in refining the representation, and selection of appropriate methods to control both external and internal processes.

- Task should be complex enough to avoid random success as much as possible. In order test not to lose its predictive power success should be safely attributed to participant's performance in solving the problem.

- Completion of the task should not require long procedures. If efficiency would be a measure of success, then the task should be quickly completed after the device model is fully understood. This would ensure that the ratio of time spent on problem solving to time spent on keystrokes is huge and determined by efficiency in problem solving activity to a great extent, rather than execution – evaluation loops.

Considering these criteria, among many others, one problem situation was chosen to be developed as an apparatus test.

Task consisted of reproducing a pattern of shapes shown to participants so that the pattern displayed in the interface screen exactly matches the goal pattern. The interface elements were a display and five push buttons. Three of the buttons were located under the screen, each coupled with a small display, and one button positioned on the right, labeled with an arrow pointing towards the screen (redraw button). An auxiliary button labeled "tamam" was positioned between the pattern card and screen. By pushing that button participants would be able to declare that the task was successfully completed (see Figure 4-11).

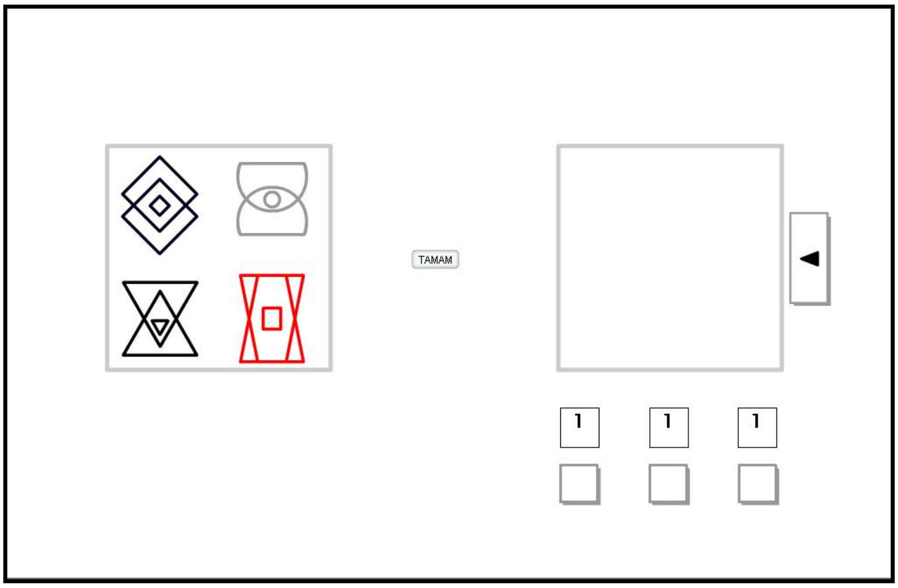


Figure 4-11 Layout of the apparatus, GIE_PS

Parameters that can be modified were not described to participants. These were as follows: (1) *slot* numbers determining where the shape will be positioned, (2) the *type* of shape, (3) and finally the *color* of the shape to be drawn. Each parameter was associated with one of the pushbuttons located under the screen. With the help of small display elements located over the pushbuttons, participants were able to see the current values assigned to parameters.

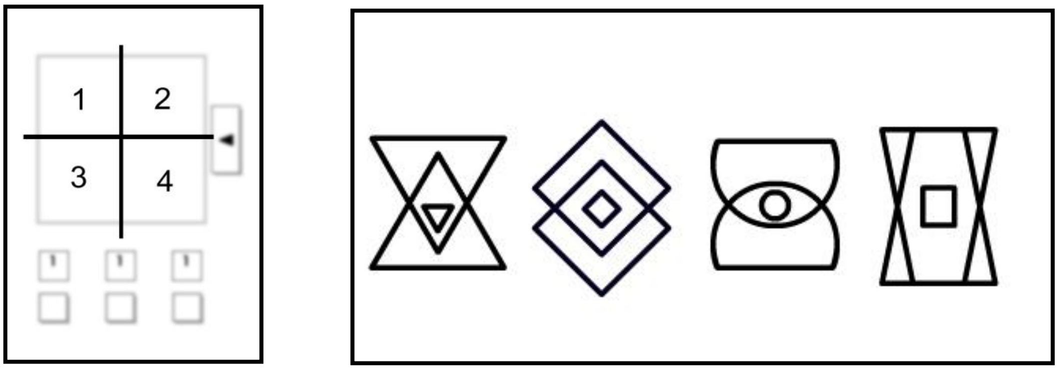


Figure 4-12 Slot numbers (left) and the types of shapes (right).

At the beginning of the test, the aim of the test was briefly described to the participants, together with some instructions about the task:

- Kullanacağınız ikinci arayüz kullanıcıların ilk kez karşılaştıkları bir ürünü incelerken geliştirdikleri yaklaşımları araştırmayı hedeflemektedir. Arayüz bir tekstil baskı makinasının sadeleştirilmiş halidir.

- Arayüz ilk bakışta kullanıcıya fazla bilgi vermemekte, çalışma mantığı ancak bir araştırma - inceleme sürecinden sonra anlaşılmaya başlanabilmektedir. Bu nedenle ilk denemelerde zorlanmanız doğaldır.

- Çalışma sırasında doğal davranışlarınızın saptanabilmesi önemli olduğundan başladığınız işlemi sonuna kadar kesintisiz ve en kısa yoldan tamamlamaya çalışınız. Sağlıklı veri toplanabilmesi için deneme bitene kadar lütfen gözlemciye soru sormayınız ve konuşmayınız.

- Arayüz fare yardımıyla kullanılmaktadır.

Amaç ekranın sol tarafında yer alan görüntünün aynısının (şekiller, renkler ver yerleşim aynı olmalı) sağda yer alan ekranda oluşturulmasıdır. İşlemin gerçekleştirilebilmesi için 4 adet tuş, 3 adet küçük gösterge ve 1 adet örnek desen ekranı kullanılmaktadır. Bunlar dışında, şekilleri fareyle sürüklemenin, şekillere ya da boşlukara tıklamanın veya klavyede herhangi bir tuşa basmanın kullanım açısından herhangi bir etkisi yoktur.

Figure 4-13 Sample Instructions form

A typical sequence of actions taken by an expert user for accomplishing the task would be as follows:

(1) Select the *slot* to be filled (see Figure 4-12) with the leftmost button,

(2) Modify the *type* parameter with the middle button,

(3) Select the appropriate value for the *color* parameter with the rightmost button,

(4) Press redraw button to see the results,

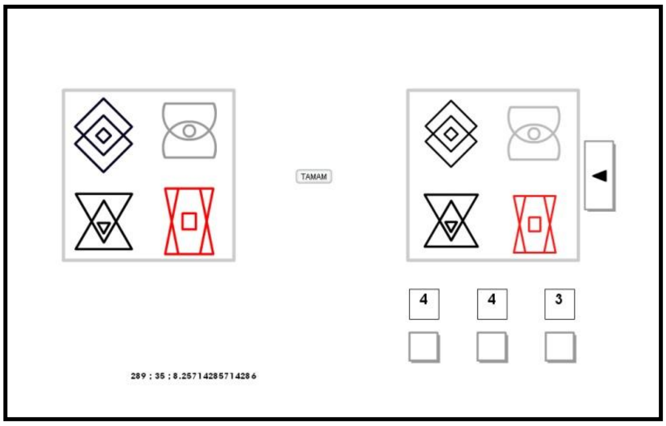


Figure 4-14 The final state

(5) After the goal state is reached (see Figure 4-14), press the button labeled "tamam".

The apparatus was modeled with Flash MX 2004, administered with a laptop PC, and participants manipulated the interface with a mouse.

After the test was implemented, a pilot study with 4 participants was conducted in order to see whether there are any technical problems.

4.1.4. Study III

Method

For gaining insight about the predictive validities of GIE_XEC and GIE_PS, tests were conducted in accordance with a comparative usability test. In that project, the aim was to comparatively evaluate four washing machines with digital interfaces. With this purpose 24 participants were allocated to three test groups, where each individual interacted with two different interfaces. The test design was as follows:

Table 4-8 Test design

Group I	Group II	Group III
Product A & Product B	Product B & Product C	Product C & Product D
N = 8	N = 8	N = 8

At the end, due to the overlapping test design, Product A and D were tested by 8 participants, where Product B and C were used by 16.

Two apparatus tests were administered to each participant[8], just before or right after the usability test sessions. Whether participants took the test before or after the sessions was not a controlled factor and was determined mainly by the restrictions imposed by test conditions.

The method of collecting the data to represent user performance was effectiveness across seven tasks. Partial effectiveness scoring was avoided since an objective way of determining partial scores seems to be impossible. Therefore, in cases where participants could not totally complete the tasks as they are defined, effectiveness was scored as 0. For each apparatus test, elapsed time data were used to represent success.

Results and discussion

Findings indicate that both GIE_XEC and GIE_PS scores correlate highly with effectiveness scores. Table 4-9 summarizes the correlation values yielded.

[8] 5 participants were not tested. Missing data will be completed and included in analyses that will be discussed during presentation of this report.

Table 4-9 Pearson's product-moment correlation between effectiveness and test scores for each product

Products	Apparatus tests	
	GIE_XEC	GIE_PS
A	-0,30	-0,95
B	-0,63	-0,39
C	-0,73	0,07
D	-0,56	-0,77

It should be noted that 6 of the participants was not successful in completing the task given in GIE_PS. Except the correlation between Product C's effectiveness and GIE_PS scores, all other values are high enough to indicate a predictive power. It should be noted that Product C had a significantly different interface design as compared to others. Whether this created a difference in correlation values is hard to tell at the moment.

If scores observed at two tests for each participant are combined, so that differences between distributions of effectiveness scores of separate tests are eliminated by converting raw scores to z-scores, the correlation between combined effectiveness and GIE_XEC was observed to be -0.70 (see Figure 4-15).

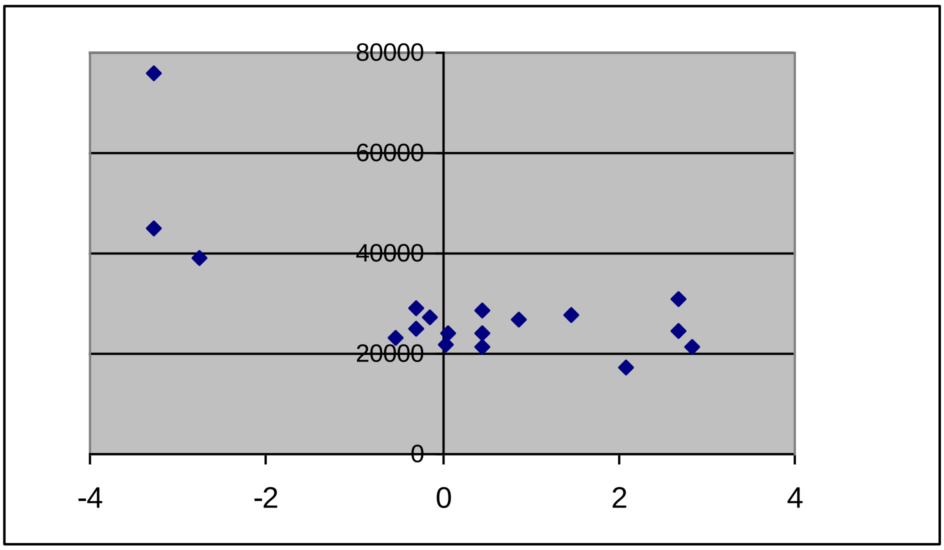


Figure 4-15 Scatter plot – Combined normalized effectiveness vs. GIE_XEC

The scatter plot of the effectiveness vs. GIE_XEC values show that there may be a non-linear relationship between two variables. If this is a valid argument, then it may be concluded that as mean time required to complete GIE_XEC increases discriminatory power of the test increases. GIE_PS, on the other hand, has yielded a correlation of -0.40.

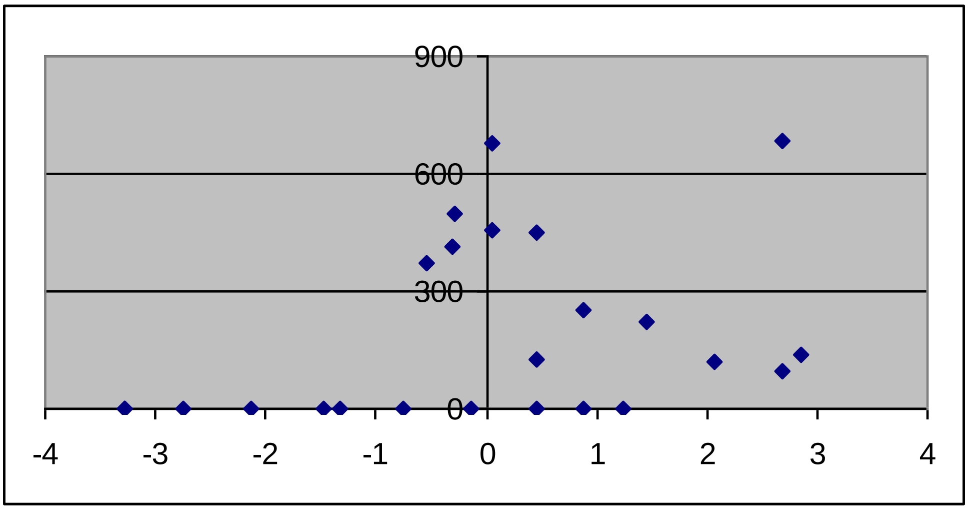


Figure 4-16 Scatter plot – Combined normalized effectiveness vs. GIE_PS

Even though this value is low, if the outlier seen on Figure 10 is eliminated this value raises up to -0,76.

The correlation between the two apparatus tests was 0,08. This result may have two reasons: (1) Since there are 6 unsuccessful participants, as opposed to GIE_XEC, GIE_PS loses its discriminatory power as GIE levels decrease. If this is true, then item difficulty should be rearranged to accommodate low GIE participants as well. (2) Results may indicate that although each test is helpful in predicting GIE levels of participants, or in other words, is correlated with success in a usability test they seem to be related with different aspects of the phenomenon. Although this explanation is in line with the theoretical assumption that types of behaviors observed in two tests are quite different, further investigations are necessary.

Considering the models of interaction presented here, types of behavior observed during interaction may be grouped under two sub-mechanisms. First group manifests itself in automatic execution – evaluation loops whereas, second group is observed in problem-solving type activities. Therefore, this dichotomy will form the theoretical foundation that justifies the existence of two separate apparatus tests. However, whether this dichotomy is sufficient to explain individual differences regarding GIE should be investigated. In the usability tests done in accordance with two apparatus tests, results indicate a high inferential power. These findings should be justified with further studies.

CHAPTER 5

5. GENERAL INTERACTION SELF EFFICACY SCALE (GISE-S)

In the following sections, first a procedure for scale development will be presented that was compiled by examining a relevant set of oft-cited scale development procedures for various purposes from the literature of psychometrics and marketing research. This procedure consists of the basic steps to follow, issues to be considered in each step, and conditions to be fulfilled in order to advance forward through the process.

In the later sections, stages of data collection will be presented, followed by successive steps of item reduction until the final form of GISE-S is obtained. In the last section, validity studies will be presented.

5.1. The characteristics of paper-based component

Many paper-based data collection techniques may be grouped under the generic term psychological tests. According to Anastasi and Urbina (1997), these range from the recognition of individuals with severe psychological and even neurological disorders to selection of personnel and "providing measures of affective variables" (4). Although, all these instruments may be accurately called psychological tests, they are dissimilar with respect to a multitude of aspects, such as their purposes of utilization, ways of development, and consequences of employing them.

According to Aiken (2000), certain dichotomies are helpful in classifying what type of instruments can be grouped under the term psychological tests. In the following lines some[9] of these classifications, provided by Aiken, that are thought to be helpful in determining the characteristics of the paper-based component, will be briefly explained.

5.1.1. Cognitive vs. affective

This dichotomy is probably the most fundamental way of classifying tests. Cognitive tests are meant to measure "the processes and products of mental activity" (Aiken, 2000), whereas affective tests assess interests, attitudes, behaviors, motives, moods, and traits. Cognitive tests may be further classified into groups such as achievement tests and aptitude tests but since such distinctions are somewhat theoretically problematic, psychologists prefer the term *ability tests* to cover the whole spectrum.

[9] *Individual vs. group* and *power vs. speed* categories were not discussed here since no decisions are necessary regarding these dimensions.

5.1.2. Verbal vs. performance

Tests may involve verbal tasks that employ entities such as diagrams and sentences or may ask respondents to perform a certain tasks like manipulating objects, sorting pictures, etc.

5.1.3. Standardized vs. non-standardized

Standardized tests are developed and administered to a large sample that is representative of the intended group and have the desired level of psychometrics properties. Often norms are developed for these types of tests. Such tests are also characterized by fixed conditions for both administration and scoring. Non-standardized tests are haphazardly brought together to fulfill an informal measurement task, such as informal course tests prepared by instructors.

5.1.4. Objective vs. nonobjective

With this dichotomy tests are classified in accordance with the strictness of the method employed in scoring. In the case of objective tests rater has no role in scoring and no special training is necessary. However, nonobjective tests are marked by the influence of raters on test scores. Certain personality tests and all essay tests are scored subjectively. However, it should be noted that objectivity concept is not used to describe the method of data collection.

After the preliminary efforts[10] to formulate the paper-based component of GIE tool and preliminary research within the related literature, it was not possible to devise an appropriate way of studying GIE with a paper-based instrument that consists of items that would spot indications of GIE. The first alternative considered was to devise a *cognitive test.* The test would be composed of items that are *verbal tasks,* where participants are asked to choose the correct action for arriving at a desired state, with a diagrammatically presented interface (see Figure 5-1).

After some items were generated it was evident that there were some serious limitations with such an approach. In *cognitive test* approach*,* scores represent the correct answers provided by subjects. Although there are cases where the degree of correctness of the answers provided may be evaluated (Nunnally, 1978), forming a causal relationship between the number of correct answers provided and subject's level of cognitive trait that is tried to be measured is indispensable. It is evident that preparation of items suitable for such an assessment is only possible when the task is overtly simple. Even there may be disputes about whether it is well-grounded to assert that c is the correct answer for the task presented in Figure 5-1. Obviously, regardless of the complexity of the problem, number of plausible solutions is almost infinite.

[10] Reported in Thesis Proposal and Report 1.

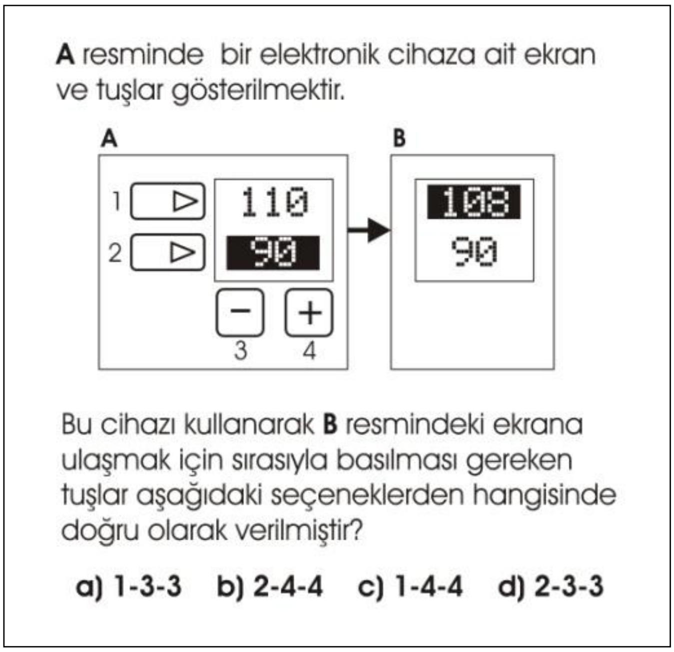


Figure 5-1 An item for a cognitive – verbal test

As the interaction task gets more complex, the severity of the problem further increases as to render such an approach totally content and face-invalid. If it was decided that including only the basic interaction tasks will alleviate the problem, items would start to loose their representative power. In other words, if only low difficulty items were included the test would only identify subjects with very low levels of GIE, and consequently loose all its predictive validity (see Figure 5-2)

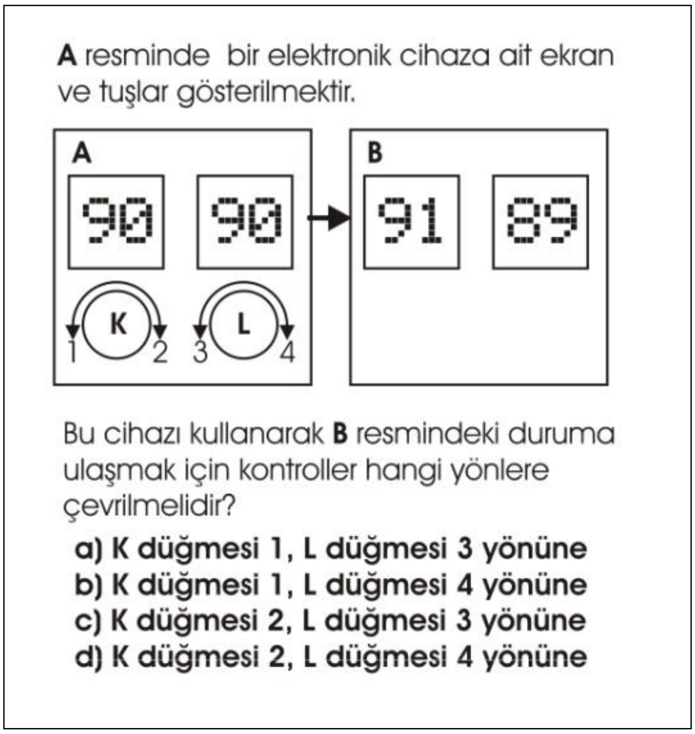


Figure 5-2 An easy interaction task formatted as a paper-based verbal item

The interaction task given in Figure 5.2 is a simple one. It may be legitimately argued that even individuals with low levels of GIE perform such tasks during their daily experience with products. However, it may not be the case for the paper-based task, which is an abstract representation of the interaction task. Therefore, apart from the fact that it is rather problematic to design interaction tasks with a unique correct solution, medium of representation brings another serious problem forward. The formal and abstract quality of the language[11] inevitably[12] used to

[11] Both visual and literal language

reconstruct the interaction experience and explain the goal state to be arrived at is likely to influence item difficulty to a great extent. In other words, the probability of a subject to successfully solve the interaction task is not determined only by subject's GIE. Most probably such a test would measure both GIE and a confounding variable, which is related to ability to decode formal notation. This would be to contaminate the scores obtained with a persistent source of serious systematic error.

Another problem with *cognitive verbal* tasks may be experienced related to the face validity of the instrument. As the tasks get easier and become more disconnected from real-life interaction, items become similar in format to that of an "IQ test". Although consisted of real-life-like tasks, this problem was witnessed even with apparatus tests and one of the participants reported that she felt like a guinea pig, being "intelligence tested". A final problem that surfaces is the instrument reactivity, that is, the subject's style of behavior may be temporarily influenced by the measurement instrument itself. After coming across with "rules of interaction" embedded in the atomic test tasks, it is likely that participants exhibit a more conservative style of interaction in a usability test conducted just after administering the instrument, with the idea that there are 'correct' ways of accomplishing certain tasks. This, in the eyes of the participants, would hinder the idea that the only purpose of conducting a usability test is to test the interface.

Having put all these, it is better to consider the alternative to specify the instrument as an *affective test* composed of *verbal* items, formulated without the use of formal/symbolic language. Decisions related to the other dichotomies are relatively easier. In order the instrument to be a sound alternative to apparatus

---

[12] A *cognitive test* item format where such formal language is avoided is impossible to devise unless the test medium is a concrete interface, as in the case of apparatus tests.

tests, ease of administration should be guaranteed. Otherwise, the virtue of developing another method would be limited to triangulation purposes. However, in practice, efficiency of administration may determine whether the instrument would be successfully employed by usability researchers and interface designers or not. Therefore, the instrument should be *objective* and suitable to be self-administered in either individual or group settings. Finally, to arrive at a *standardized* test is the ultimate goal of this project. However, whether it will be possible to attain the level of refinement necessary for the instrument to comply with the criteria is hard to tell at the moment.

5.1.5. 'Scale' as an alternative to cognitive test

By considering the specifications for the instrument, coarsely put above, it can be stated that measurement scales are appropriate for the measurement task.

Measurement scales are widely used instruments developed and administered to measure various constructs in social sciences (Spector, 1992) and marketing research.

Apart from their similarities with ability tests, scales rely on *sentiments*, which are responses given without any *veridical comparisons,* where correct judgments are attributed to the skill/ability under scrutiny (Nunnally, 1978). The constructs targeted by scales are mostly psychological entities such as personal interests, attitudes, and beliefs. Therefore, if coarsely put, by utilizing a scale, the researcher aims to measure a construct with the use of self-reported data provided by respondents. Nunnally formulates this major distinction accurately as follows:

> In the scaling of people, all tests of ability concern judgments, in a broad sense of the term. This is true in tests of mathematics, vocabulary, and reasoning ability. The subject either exercises judgment in supplying correct answer for each item or judges which of a number of alternative responses is most correct[…]Measures of attitudes and personality can require either judgments or expressions of sentiment[…] One can make a good argument for referring to judgment as concerning "knowing" and sentiments as concerning "feeling".
>
> (43)

Consequently, by deciding that a measurement scale will be developed, one not only expresses that there is an intention of measuring a variable but also how that variable is approached epistemologically.

For example, one can attempt to measure ability to solve algebraic problems with a set of items that contain problems sampled from the domain of algebra. If this is the case, the number of items answered correctly would be an accurate indicator of subject's ability to solve problems of this sort, since subject's problem solving performance is somehow quantified and the instrument may be considered 'objective' in this sense. However, if one attempts to measure people's attitude towards algebra there is no 'objective' way of quantifying this trait.

5.2. The concept of 'latent traits/constructs'

As defined by Cronbach and Meehl (1955), a construct is an attribute postulated to be possessed by individuals and reflected in behavior (as 'test performance' in their context). It is designed to be utilized in a scientific study, "generally to organize knowledge and direct research in an attempt to describe or explain some aspect of nature" (Peter, 1981). It is only possible to make inferences about the attribute by examining its surface manifestations. Therefore, constructs can be observed indirectly. However, if a construct cannot be observed at all then it is just a metaphysical entity (Peter, 1981).

In the algebra test example given above, the construct that is being investigated was "ability to solve algebraic problems"—i.e. ability to solve problems that are similar to the ones included in the instrument. However, if the construct is defined as "algebraic ability" then, it is not possible to improvise an instrument. An alternative model of measurement called *latent trait models* are founded on this basic idea that constructs can only be studied by examining their indicators:

> (1)There must be a stimulus variable, or set of a variable, that is presented to individuals. These variables can be, for example, test items on an ability test or an achievement test, personality questionnaire items, or items on an attitude scale.
>
> (2)The items are presented to an individual, and they elicit certain responses that are observed and recorded.
>
> (3)To enable the psychometrician to infer a person's status on the trait based on the observed responses to a specified stimulus variable, or set of stimulus variables, the hypothesized relationships between the observed responses and the underlying trait levels are formalized by an equation that describes the functional form of that relationship.
>
> (Weiss, 1983, p. 1)

Consequently, having decided that the instrument should be an *affective* one, the construct[13] to be measured may be conceptualized within a latent trait model.

[13] A construct that is to be defined in the theoretical vicinity of GIE

Thus, development procedure should commence with how this latent construct can be defined and what may be the types of responses associated with it.

5.2.1. 'Reflective' and 'formative' measures for constructs

According to Netemeyer, Bearden and Sharma (2003), manifestations associated with the construct to be quantified may either be formative or reflective. If an instrument relies on formative measures of a construct, then this instrument may be called an index, not a scale. If the instrument is an index, items 'form' that construct, in other words, items may ask subjects to give information about factors that are thought to cause the construct (see Figure 5-3).

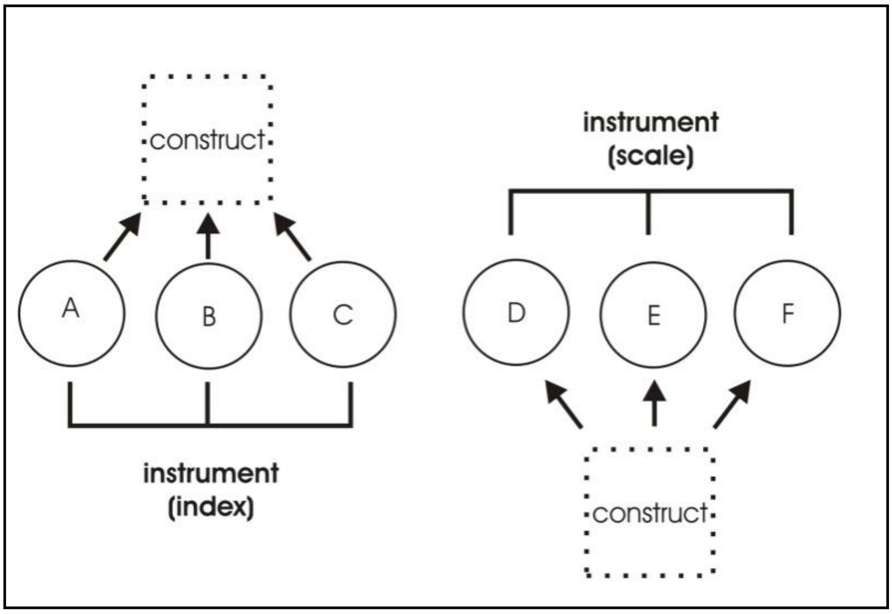


Figure 5-3 Formative and reflective measures

Therefore, magnitudes of formative indicators (A, B, C in Figure 5-3) determine the magnitude of the construct. However, magnitude of the construct does not affect each indicator (Diamantopoulos and Winklhofer, 2001). Index of socioeconomic status (SES) is a widely used mechanism to illustrate the relationship between formative indicators and constructs (see MacCallum and Browne, 1993). As indicators of SES (income, education level, occupation and residence) increase SES also increases, but if SES increases this is not reflected to all indicators.

In the case of reflective measures, indicators (D, E, F in see Figure 5-3) reflect the level of construct. Therefore, each indicator is an individual variable that correlates with the magnitude of trait to be measured.

In the case of GIE, in order to propose an instrument that relies on cause indicators, more theoretical elaboration on the causes of GIE is necessary. Therefore, focusing on reflective measures seems to be the appropriate choice at the moment. Besides lack of a theory on causes of GIE, techniques for developing instruments based on reflective measures are wide-spread and well-developed.

5.3. Scale development procedure

Before taking any further steps for construct definition and identification of responses, a concrete scale development procedure should be adopted. In this section the literature review done for compiling an appropriate procedure will be presented.

Scale development is a broad subject area covering methodology related domains of many disciplines such as psychology, sociology, marketing, organizational behavior, personnel selection, and ergonomics[14].

In order to identify the essential steps that will form the basic structure of procedure, both basic material on fundamentals of scale development (e.g. DeVellis, 1991; Netemeyer, Bearden and Sharma, 2003; Churchill, 1979; and focused discussions on technical and theoretical issues were reviewed.

[14] Unlike ability tests, scaling instruments are utilized in a diversity of contexts where measurement of a latent construct is necessary.

After the comparative examination of the selected procedures, some attributes that are common in all of them were identified. Almost all the procedures comprised of detailed descriptions of concrete steps to be taken for a satisfactory scale. The main procedures were usually accompanied with easy to follow techniques, so that what should be done in each step was clearly defined with operational suggestions and examples. Although most of the procedures were represented as sequential processes, the iterative nature of the development task was usually emphasized. After reviewing the selected literature, it was apparent that, maybe the most critical aspect of development is to decide where to terminate the iterations. Another common strategy employed by all the examples was to 'construct' the scale in an inductive fashion. As a consequence of this strategy, suggested procedures were easy to analyze into two main stages, namely theoretical and empirical phases. It was recommended that the research should start with a thorough theoretical study, so that both existing theories are judged in terms of their suitability to define the construct and new models may be proposed where the existing ones cannot cover the research area extensively. Subsequently, items that are thought to be useful for scaling the construct delineated in the theoretical phase are tested empirically. Until the desired level of reciprocity and item quality is attained, items are refined. Although not cited within the basic material, there are some studies suggesting that the development process should be lead by empirical findings, which is called criterion-keying. According to this view, first, researcher should go through the empirical phase and show deductively that certain items from a variety of theoretical origins are useful in predicting a certain behavior, which is closely related with the construct to be measured. However, such a strategy is not easy to follow in the present case. Even if some serious problems concerning reliability[15] are ignored, the fact that behavior to be predicted should certainly be usability test performance makes it impossible to work with a large sample as far as the extent of resources to be allocated in the study are considered. Furthermore, some theoretical models inclusive enough for constructing a definition for GIE are present.

In Figure 5-4, the main steps of the procedure compiled as a result of this comparative analysis are presented.

[15] These will be briefly pointed out in the following sections.

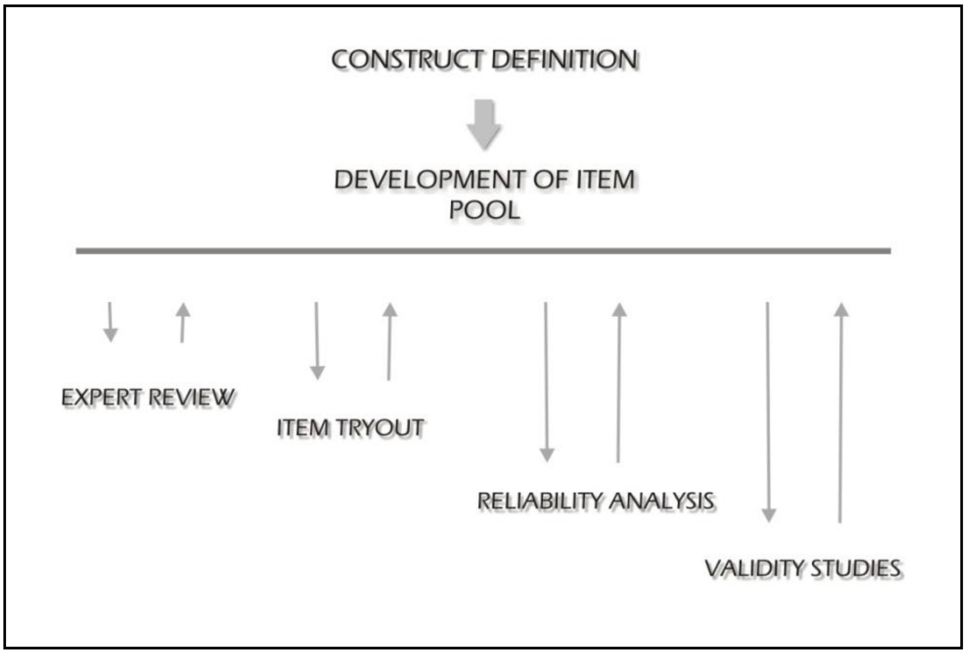


Figure 5-4 Main steps in scale development

As it is obviously apparent, the procedure 'proposed' here actually consists of steps and basic structure that underlie the models compared. Therefore the procedure may be considered as the resultant structure arrived at by collapsing the models into a single procedure.

Before a detailed description of each step and converting this structure to a working algorithm, some implications of adopting such a procedure should be listed. First of all, before any major data collection, there is one semi-empirical step where expert view is consulted and an item tryout step, which may be considered as a pilot study focusing on item characteristics. These two preliminary steps are followed by two sessions of major data collection, former concentrating on item reliability and the latter on whether the instrument measures what it ought to measure.

It should be noted that, after each step, item pool is refined by removing bad items and introducing new items if necessary. It may be necessary to revise the construct definition and the general characteristics of item pool in the case that instrument is not properly validated. Some additional steps may be included in order to check for predictive validity with the item pool at hand if any opportunities for usability tests arise.

5.3.1. Step 1: Construct definition

Construct definition is considered a crucially important step often overlooked in scale development, since a well conceptualized construct is essential for a valid instrument to be developed. What is worse, failure at this step may be hard to notice before validity studies, which means invaluable investment of resources will still be made up to that step (DeVellis, 1991). A clear definition may be very helpful while generating items (Spector, 1992) and initial judgments of item appropriateness can be based on benchmarking each item against this definition.

According to Netemeyer, Bearden and Sharma (2003), an important dimension to consider is the scope of the construct. If the scope is too narrowly defined then some important facets of the construct could be missed. This is referred to as *construct under representation* and may hinder both reliability and validity of the instrument. At the other extreme construct definition may be too broad so that items generated in accordance would measure other constructs as well.

Consequently, *construct-irrelevant variance* is introduced as a systematic source of error. Furthermore, if more than one variable is being measured than problem of content heterogeneity arises. This problem is accurately delineated by Smith and McCarthy (1995). They argue that if a scale's contents bear too much resemblance to another scale that measures some similar but different construct, an illusive situation is confronted with.

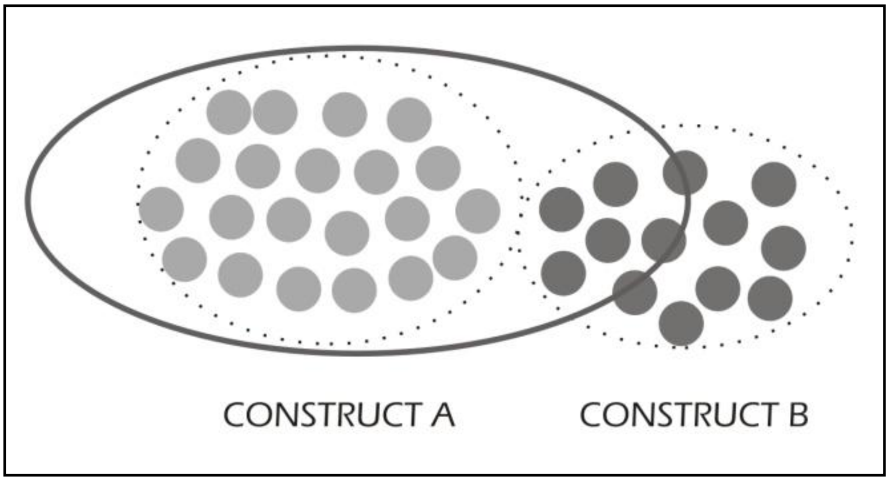


Figure 5-5 Content heterogeneity

If a construct is broadly defined, crosscuts and intersections with proximal constructs are inevitable. Consequently, items that fall within the scope of the construct can co-exist in the domain of another scale (see Figure 5.5). Under such circumstances, the scores obtained with these scales will be attenuated, not as a function of a causal relationship in between but as a function of the area of intersection between two constructs. However, it should be noted that it is not a mistake to define a broad scope for a construct unless its consequences are known. The dotted regions depicted in Figure 5.5 should not be regarded as 'real' boundaries of constructs, since boundaries are 'constructed' not 'discovered'. The problem here is to mistake the effects of a confounding variable for an indication of causal relationship.

In order to overcome problems of this sort, Cronbach and Meelh's (1955) early concept of *nomological network* is useful. As far as a construct is defined within a network of other constructs in the vicinity such problems are not likely to be experienced.

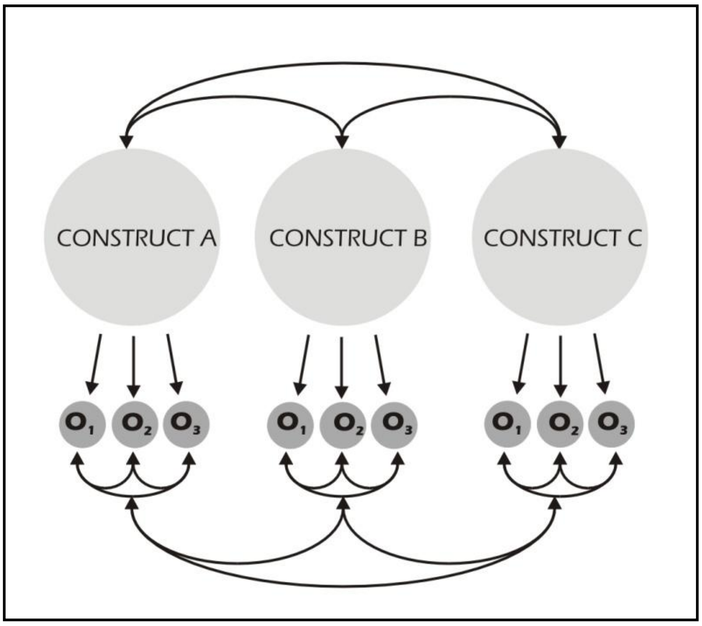


Figure 5-6 Nomological network [16]

Some of the principles of the nomological net may be enumerated as follows[17]:

- The nomological network is an interlocking system of laws
- These laws may specify the relations shown in

[16] Adapted from *The nomological network*, online document http://www.socialresearchmethods.net/kb/nomonet.htm, retrieved in August 12, 2006

[17] see Cronbach & Meehl (1955) for the complete set of principles

- **Figure** 5-6—i.e. relationship between constructs, between constructs and observables, and between observables.
- A construct may only be scientifically defined if it is defined in a nomological network.
- If the nomological network is elaborated the knowledge about a theoretical construct increases.

These basic principles indicate that it is not possible to define a construct in isolation. Therefore, what is excluded from a construct is just important as what is included (Churchill, 1979; Clark and Watson, 1995).

In this step for deciding on the entities to be included and excluded, literature research plays an important role in identifying and studying "previous attempts to conceptualize and assess both the same construct and closely related constructs" (Clark and Watson,1995). Finally a brief, unambiguous operational definition that reflects the essentials and all the facets of the construct should be provided. However, after iterations, whether this tentative definition should be checked and refinements or revisions are necessary should be considered.

5.3.2. Step 2: Development of item pool

Having arrived at an operational definition of construct, concrete formulations for data collection—i.e. generation of items—should be handled at this step. At this point it should be remembered that first departures from the construct are witnessed as well. Put in a different way, since there are no ideal items that overlap with construct definition perfectly, the instrument unavoidably starts to lose its pertinence and error components contaminate the process. The aim should be to employ strategies that will minimize the infiltration of 'impurities' to the item wordings. It should be noted that the qualities of items in fact determine whether the construct is situated accurately within the network of constructs and not the construct definition on its own.

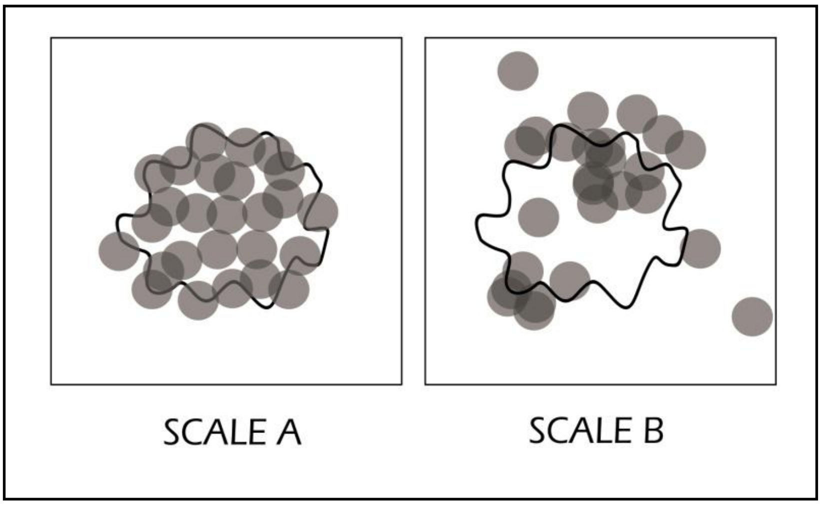


Figure 5-7 Good and bad item distribution

The ultimate role of the quality of item pool is depicted in Figure 5.7. Although both scales have a common construct definition, items in scale b have poor item distribution properties regarding both homogeneity of distribution and accuracy of item positioning.

On the other hand, item pool for Scale A is so accurate and homogenously distributed that there are almost no items that are off the target or overlap with other items. Of course, in reality, items do overlap more and this is not always an indication of poor item quality. The relation between redundancy and reliability will be discussed later in this report.

Although item writing is a step to be handled with utmost care there are neither straightforward analytical techniques for item writing (Clark and Watson, 1995), nor guaranteed-to-work methods of monitoring item quality. This step in scale development is often called an art rather than science.

Up to now, the main focus of the discussion was related with the success in theoretical elaborations of the construct and writing items that sample that domain well. However, respondents who provide responses to the items also undergo a complex cognitive process, which may be a serious error source in itself. Krosnick, Judd and Wittenbrink (2005) state that the process is comprised of three stages: a) activation of memory contents after reading the item, b) deliberation on the contents of memory, and finally c) a response (p. 24). Tourangeau and Rasinski (1988) describe the process and its outcomes as follows:

> Respondents first interpret the attitude question, determining what attitude the question is about. They then retrieve relevant beliefs and feelings. Next, they apply these beliefs and feelings in rendering the appropriate judgment. Finally, they use this judgment to select a response. (p. 299, also qtd. in Oskamp, 2004)

There are three junctions in the process where certain transformations and loss of accuracy may occur. If this three-step process is integrated to the measurement model previously suggested, the number of critical junctions in the whole process increases (see Figure 5-8).

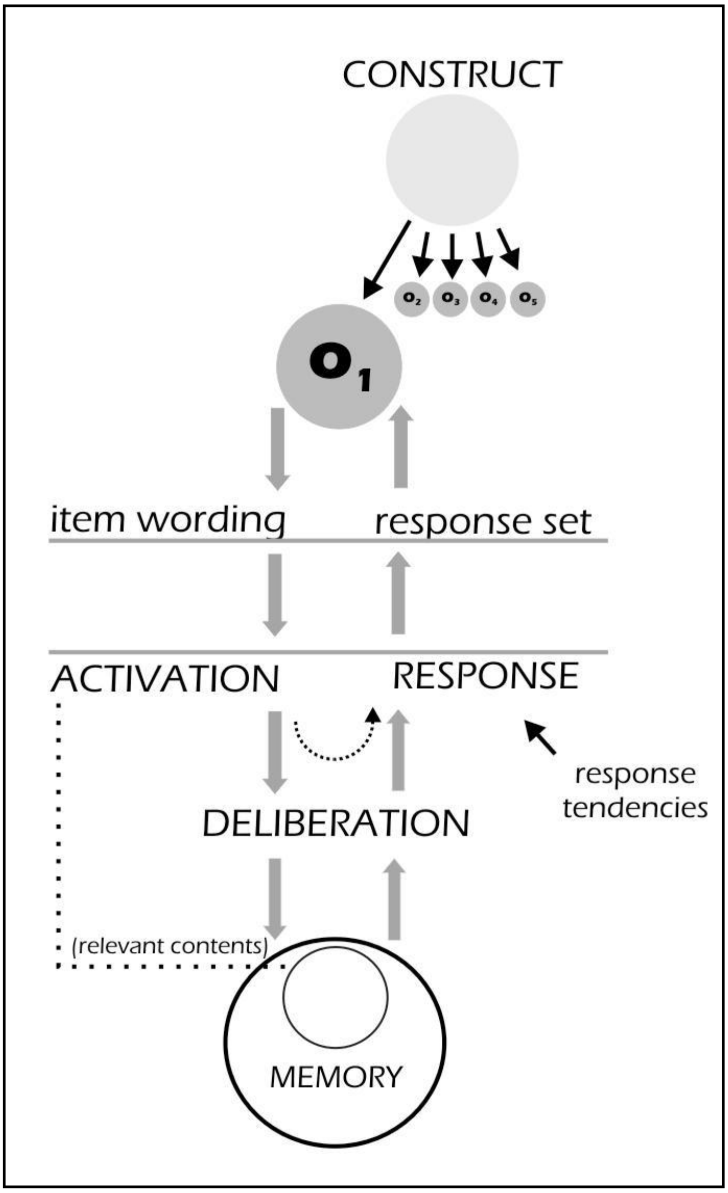


Figure 5-8 Process of providing response

In the following lines, this process will be investigated considering the sources of problems specific to each transformation.

Item wording ↔ activation

As suggested before, item wording utilized as a stimulus is expected to induce a certain activation of the related memory content. However, inaccurate wording can lead to confusions and consequently the memory content retrieved may be irrelevant. Common sources of such error are enumerated below:

- Use of colloquialism or jargon
- Long items
- Double barreled items
- Double negatives
- Items with weak statements (a problem specific to items that employ Likert scale)

(e.g. Churchill, 1979; :DeVellis, 1991; Spector, 1992; Netemeyer, Bearden and Sharma, 2003)

Deliberation ↔ memory content

There may be items that ask for attitudes, feelings and beliefs that respondents have no pre-established idea (Krosnick, Judd and Wittenbrink, 2005). Inclusion of such items may jeopardize the psychometric qualities seriously.

Oskamp states that this problem arises when respondents improvise and provide an answer on spot.

> [T]he fact that people sometimes *construct* attitude responses on the spot without any prior consideration of the issue, rather than retrieving a previously formed attitude from their memory, would sharply decrease both the reliability and validity of such attitude statements.
>
> (Oskamp, 2004, p. 57)

Following examples may be helpful in illustrating the problematic nature of such formulations:

- Cep bilgisayarlarını kullanmakta çok zorlanırım[18] (I will have a hardtime while using a pda)
- Connect 4510 çok rahat öğrenilen bir telefon (Connect 4510 is an easy-to-learn phone)
- Yeni aldığım cep telefonunun kullanımı eskisinden farklıysa çok sıkıntı çekerim (If the new phone I buy has a different style of use I will suffer much)

For a respondent to answer the first item a quite specific type of experience is necessary. It is quite likely that a majority of respondents would not be able give a

[18] For examples to provide guidance during item generation and refinement, they are structured in Turkish.

response depending on a previously established attitude. In the second item, again a specific experience is asked for, but this time probably item is going to lose its meaning after the product that is referred to becomes obsolete. In the last example the subject is asked to report her/his typical feelings in a rarely occurring event. The common problem observed with these examples is that subjects are forced to make speculations on issues without any relevant memory content.

Another problem witnessed in this stage is the 'item difficulty' as it is called in the literature of classical ability testing. Items should not include statements that will be endorsed or negated by a very large portion of the respondents (e.g. Clark and Watson, 1995). Although they may be validly situated within the construct defined, such items have no differentiating power, and therefore should be discarded.

Deliberation ↔ response

There may be cases where the outcomes of the deliberation are influenced by some other external factor. Other global response tendencies, strategies or lack of cognitive resources may influence the responses given. Johnson (2004) states that especially how people *perform* in social life, in order to portray a profile, has a determining effect on their style of responding to questionnaires or scales. In other words, responding to items of questionnaires cannot be considered separately from other social activities. Adopting a similar approach, Hogan (1991) argues that responses to items are "automatic and often nonconscious efforts on the part of test-takers to negotiate an identity with an anonymous interviewer (the test author)" (p.902, also qtd. in Johnson, 2004)[19]. Within a *constative* perspective, Oskamp lists the factors that influence responses and are external with regards to the construct investigated as follows:

Carelessness – respondents may show low motivation to fill out the scale. Although appropriate instructions, reducing item length and limiting number of items may help to alleviate the problem, all the forms should be scanned for obvious indications of careless responding, such as many left-out items, pattern filling, etc.

Social desirability – This phenomenon is witnessed when respondents give answers in order to be on the socially desirable side or to conform with the cultural norms (Netemeyer, Bearden and Sharma, 2003). Nonetheless, in the case of GIE, which is planned to be applied in contexts where no performance assessment or selection is done, social desirability may not pose a serious problem compared to, for instance, any instances of personality research. However, particular care should be exercised to neutralize the effects of social desirability bias if such items are recognized.

Acquiescence – Respondents may show the general tendency to endorse items regardless of the statement embedded in the item stem. It is a recommended

[19] Johnson, in his article *The impact of item characteristics on item and scale validity*, offers a critical look to the mainstream approach (constative approach) that assumes respondents retrieve memory contents when prompted and there may be 'poor' item characteristics that may deviate their answers. The 'performative' approach, as an alternative view, does not attest that some response patterns (such as social desirablity bias, acquiscence, etc.) do not affect validity to a great extent. Johnson provides empirical evidence that items that are easily associated with the trait to be measured influence the results with regards to validity.

Although, the approach is theoretically appealing in the sense that it considers people usually do not use language to communicate propositional statements, studies that show its merits in practice are not much. As far as this study is considered, such methodological discussions are too specific.

practice to reverse half of the items—called a balanced scale (Oskamp, 2004)—so that endorsing all the items would not yield a high total score.

According to Krosnick (1991), almost all the deviants may be associated with a behavior termed ‘satisficing’. In line with this approach, Krosnick argues that tasks with high cognitive demands, respondent’s low level of ‘cognitive sophistication’, and low motivation to respond are the conditions that stimulate satisficing. As a result, subject may choose the alternative that she/he identifies as the ‘correct’ answer, may agree with all assertions—i.e. exhibit acquiescence, accept statements maintaining *status quo*, respond all the items with the same rating on the scale, say ‘don’t know’, and exercise mental coin-flipping.

While generating the pool of items, it is recommended that, facets of the construct should be proportionately represented by the items (e.g. Smith and McCarthy, 1995; Haynes, Richard and Kubany, 1995). For aggregated measures where the sum of individual item ratings is regarded as total score, the danger of disproportionate representation is apparent.

For items to suit the purposes of the instrument and in order to ensure that the irrelevant or poorly worded items are excluded, semi-structured interviews and focus groups conducted with the target population are recommended (e.g. Churchill, 1979; Dawis, 1987; Haynes, Richard and Kubany, 1995)[20]. Since present study involves the development of an instrument to measure the competency of individuals in using digital consumer products the target population is quite

[20] In cases where the target group has its own culture it may be crucial to conduct exploratory work. For example, an instrument to measure self-perceived innovativeness being developed to assess designers will definitely necessitate collecting preparatory data that will guide both construct definition and item wording.

large[21]. Therefore, it may not be possible to detect a coherent body of beliefs, customs, and terminology interiorized by all the members of the target population.

General strategy to be followed in item generation

After revisiting some general methodological concerns in item generation, in this section some general strategies that will ensure that an item pool is suitable for further refinements in the later stages, will be presented.

All the procedures included in the comparative analysis emphasize reduction of the number of items initially generated. What is meant by item refinement is actually discarding the items that are far from attaining certain criteria. Techniques for accomplishing this subtractive task consist of keeping items that do not harm content validity, unidimensionality, reliability, and certain types of validity. These concepts and corresponding techniques will be handled in detail later throughout the development process. Here, a general strategy to ensure that there are enough items in the initial pool of items will be provided, since the success at later stages depend on the inclusiveness of the set.

Referring to Loevinger's ideas on content sampling, Clark and Watson (1995) recommend that all the content that may be included in the construct should be represented as much as possible. By doing this, researcher tries to ascertain that items do not only reflect the components of a theory initially chosen to guide the process. The benefits of this strategy are expressed by Clark and Watson (1995) as follows.

---

[21] Theoretically all the people in universe may be considered in the target population.

> Two key implications of this principle are that the initial pool (a) should be broader and more comprehensive than one's own theoretical view of the target construct and (b) should include content that ultimately will be shown to be *tangential* [emphasis added] or even unrelated to the core construct. The logic underlying this principle is simple: Subsequent psychometric analyses can identify weak, unrelated items that should be dropped from the emerging scale […]. Accordingly, in creating the item pool one always should err on the side of overinclusiveness.
>
> (p. 311)

The implications of being 'overinclusiveness' in the process of setting up the item pool are numerous, but one of them should be highlighted here. Redundancy is an inevitable consequence that is often encouraged to overcome problems with item specific errors (DeVellis, 1991). Actually, any instrument that depend on aggregated total scores obtained by employing multiple i enjoy item redundancy. However, redundancy should not be interpreted as scales should include item stems that have the same content with slight differences in wording.

Although it may sound like an atheoretical approach, it is often suggested that construct should be revised as new aspects of the trait investigated are brought to lime light by empirical studies (e.g. Smith and McCarthy, 1995). If the construct belongs to a domain that is not studied extensively it will take many attempts to accurately delineate the construct (Spector, 1992).

5.3.3. Step 3: Expert review

Expert review is listed among the techniques that aim to refine the item pool without the involvement of the target sample. Technique is based on the assessment of the items individually considering "relevance, representativeness, specifity, and clarity" (Haynes, Richard and Kubany, 1995). According to Crocker and Algina (1986), items should also be checked for technical item-construction flaws, offensiveness or bias, readability, problems, and grammatical errors.

In order the committee of experts to evaluate appropriateness of items with regards to the construct under scrutiny, a thorough definition of the construct should be provided (DeVellis, 1991) together with a brief instruction and a guideline that includes rules for good item design.

Experts may be asked to map their comments in a structured way with the use of a rating scale. The upper portion of the item set ranked after employing a scoring scheme based on the ratings provided may be kept. Furthermore, some new items, and even facets of the construct may be suggested by the experts. For the present study, experts are planned to be chosen among researchers with a considerable experience in user research.

5.3.4. Step 4: Initial item try out

After the item refinement in the light of expert review, items may be tested with a small sample of representative subjects (N = 30-50). In this step either response data, or the actual behavior of subjects while responding to items may be focused. Crocker and Algina (1986) state that gathering observational data is useful for identifying ambiguous or hard-to-respond items, by assessing the distribution of response latencies. Furthermore, descriptive statistics may be exploited for identifying further flaws:

- Response variances yielded for every item may be checked for identifying items with too high or too low item difficulty.
- Items that behave unexpectedly may be identified by checking interitem correlations.
- Response latencies may be measured for identifying poor items
- Items that cause subjects to change their minds frequently may be spotted and either re-worded or discarded.

As a complementary technique, a concise debriefing session can be held right after the subjects complete the scale. Subjects may be asked to report ambiguous wording, irrelevant content, or use of jargon. Literature should be further researched for studies that specifically discuss similar techniques and the use of descriptive statistics in item analysis.

5.4. Construct Definition

As it was discussed in Chapter 3, the concept of **'self-efficacy'** proposed by Bandura (1986) is frequently utilized to measure and even predict performance. According to Bandura, individuals possess a self system that enables them to influence their cognitive processes and actions. Therefore, "what people know, the skills they possess, or what they have previously accomplished are not always good predictors of subsequent attainments because the beliefs they hold about their capabilities powerfully influence the ways in which they will behave" (Pajares, 1997). In line with this view, researchers developed many scales that targeted 'computer self-efficacy' (e.g. Murphy, Coover and Owen, 1989; Compeau and Higgins, 1995; Quade, 2003; Barbeite and Weiss, 2004; Torkzadeh and VanDyke, 2001).

Suggested as 'more than just a mere reflection of performance', the concept of 'self-efficacy' was considered as a framework for defining the construct that will form the backbone of the scale under development.

5.4.1. Measuring self-efficacy

Before an attempt of construct definition is made things to be considered in measurement should be revised, since how the construct is defined determines how the characteristics of the instrument.

The aggregate nature of constructs such as General Computer Self-Efficacy (Marakas, Yi and Johnson, 1998) makes it quite plausible from a perspective of measurement. Marakas, Yi and Johnson (1998) describe this as follows:

> In particular, we believe that given the definition of GCSE as a collection of CSE perceptions and enactive experiences, GCSE does not intuitively appear to be amenable to a measurably immediate change under any set of short-lived conditions. Correspondingly, its long-term usefulness may be as a predictor of future levels of general performance within the diverse domain of computer related tasks.
>
> (p. 129)

Being comprehended at this level, a potential source of error, that is temporary changes in construct to be measured, is eliminated at least on theoretical grounds.

According to Compeau and Higgins (1995)[22], this holistic comprehension of the construct should be reflected to the approach adopted in measurement. It is argued that concentrating on individual sub-skills rather than self-efficacy beliefs for accomplishing tasks is a misconception exhibited by some researchers.

For example, the scale developed by Murphy, Coover and Owen (1989) aims to arrive at a compound score of computer self-efficacy by investigating atomic skills such as 'Moving the cursor around the monitor screen' or 'Calling-up a data file to view on the monitor screen'.

While discussing the common errors in assessment, Bong (2006) maintains that self-efficacy should not be confused with other self-referent constructs such as self-esteem and self-concept.

> The most common mistake is to assess self-efficacy as a domain-specific form of self-esteem. Investigators who commit this error conceptualize self-esteem as a global index of perceived self-worth spanning across many disparate domains and self-efficacy as similar emotional reactions toward the self but in specific domains. However, self-esteem need not be detached from a functional domain, nor is there a part-whole relationship between self-efficacy and self-esteem (Bandura, 1997) [ctd. in Bong 2006].
>
> (p. 289)

Therefore, constructs that claim to be a type of self-efficacy should concentrate on one's confidence in accomplishing a task, and not self-worth or self-perceptions regarding a specific domain.

---

[22] A scale that aims to measure computer self-efficacy is developed by Compeau and Higgins. Although, not the most popular scale, it is widely cited as a comprehensive attempt to define and measure computer self-efficacy. A reprint is provided in Appendix I

Another error to be avoided is stated as ignoring the context-specific and generative nature of self-efficacy constructs. Consequently, measurements should not be based on self-assessments done in vacuum and respondents should not be forced to weigh their self confidence on highly abstracted situations. Finally, Bong (2006) warns that beliefs that match what is to be predicted should be looked for. In other words, it is asserted that "the predictive utility of self-efficacy is maximized when these beliefs are estimated in reference to the tasks and contexts that best correspond to the criterial variable (Bandura, 1997; Pajares, 1996) [ctd. in Bong 2006, p.295].

Bandura (2006) in his book chapter *Guide for Constructing Self-Efficacy Scales*, states that perceived capability should be targeted by items "phrased in terms of *can do* rather than *will do"* (p.308) so that intentions are not mistaken for self-efficacy perceptions. Another crucial elaboration made by him is the danger of focusing on outcome expectancies.

> Another important distinction concerns performance outcome expectancies. Perceived self-efficacy is a judgment of capability to execute given types of performances; outcome expectations are judgments about the outcomes that are likely to flow from such performances.
>
> (p. 309)

5.4.2. Definition of the General Interaction Self-Efficacy

General Interaction Self-Efficacy (GISE) is specified as individuals' self-efficacy perceptions as far as learning new devices. Although, core definition seems to be too specifically formulated, as far as functional use of the corresponding scale is considered, both GIE and GISE are primarily utilized for predicting participant performance before usability tests are conducted. Therefore, long-term appropriation of digital products, or long-term transformations witnessed in the nature of interaction should not be engaged with as the main area of interest. However, as it was discussed in *Report 2*, it is better not to act over exclusive at this stage of instrument development.

General Interaction Self-Efficacy (GISE) is a judgment of capability to establish interaction with a new device and to adapt to novel interaction situations…

In accordance with this definition, GISE has a two-fold character. First of all, GISE is related with learning to use new devices. In this regard, it is the capability to learn how to interact under unfavorable conditions, as well as ability to sustain learning in the absence of factors that enhance the learning process. Secondly, it is the ability to reorient, recover interaction and survive in a multitude of breakdown situations. Hence, GISE targets the self-efficacy perceptions about putting GIE into use during controlled processes.

5.5. Item generation

After an initial attempt to compile a list of items that target the construct of GISE and relevant examples were examined, it was decided that a questionnaire for basing item stems on users' perceptions was necessary. Since definition of GISE has been limited so that routine interaction and long-term processes were excluded, the questionnaire targeted the early phases of coming across a new interface, and initial steps of appropriating it. The aim was to grasp the users' perceptions about factors that influence learning processes positively or negatively. The rationale behind asking users things that make learning harder or easier was to investigate whether a model could be extracted that would guide all the scale development process, as well as exploring their jargon and approach to the subject matter.

5.5.1. Methodology

Data collection was done with a self-administered questionnaire, titled *Learning Electronic Devices Questionnaire* (LEDQ), which consists of open-ended questions. The questionnaire was preceded by a one-page introduction, where aim of the study and definitions were made clear by examples (see Appendix A for a sample form). In the second part, first respondents were asked to report favorable and then unfavorable situations for learning electronic devices. LEDQ was applied both in printed and in electronic form.

Sampling was done with snowball technique. The only concern was to make sure that approximately half of the respondents were youngsters with quite strong beliefs of GISE. 102 respondents participated in the study, with an average age of 29.9 (min. 18; max. 64). 59 of the questionnaires were in printed form whereas 43 were in electronic format. Questionnaires were answered in privately. Together with the core data, age, gender, occupation and education data were asked for.

5.5.2. Results and analysis

A total of 287 negative and 269 positive expressions (550) were collected (see Appendix B for full list). Expressions were not modified as much as possible, and the main strategy was to maximize the number of potential item stems. As a result, 425 expressions were identified and an abundance of item stems with almost-redundant wordings were kept for later reduction. The data obtained were then analyzed with two main purposes. At first step, the expressions were grouped and a phenomenological model was developed (see Figure 5-9). This model was supposed to serve as a guide for ensuring content validity, and as a structured item pool. It should be noted that such a model should not be mistaken for a factual model based on empirical findings. The rationale behind constructing such a model is to gain insight about users' perceptions about learning process and having a structural representation for guiding the rest of the development process.

First order elements in the collective phenomenological model were novelty and familiarity, affection, usefulness, ease of use, help and support, learning context and process, breakdowns, and prior knowledge. Note that, as it was intended, the majority of groups were based on either traits of artifacts or of interaction, except prior knowledge. In the table below, the distribution of number of items across 8 groups was provided.

Table 5-1 Distribution of items[23]

Sub-construct	N
novelty and familiarity	42
affection	33
usefulness	35
ease of use	138
help and support	119
learning context and process	33
breakdowns	15
prior knowledge	10

[23] See Appendix C for expressions included.

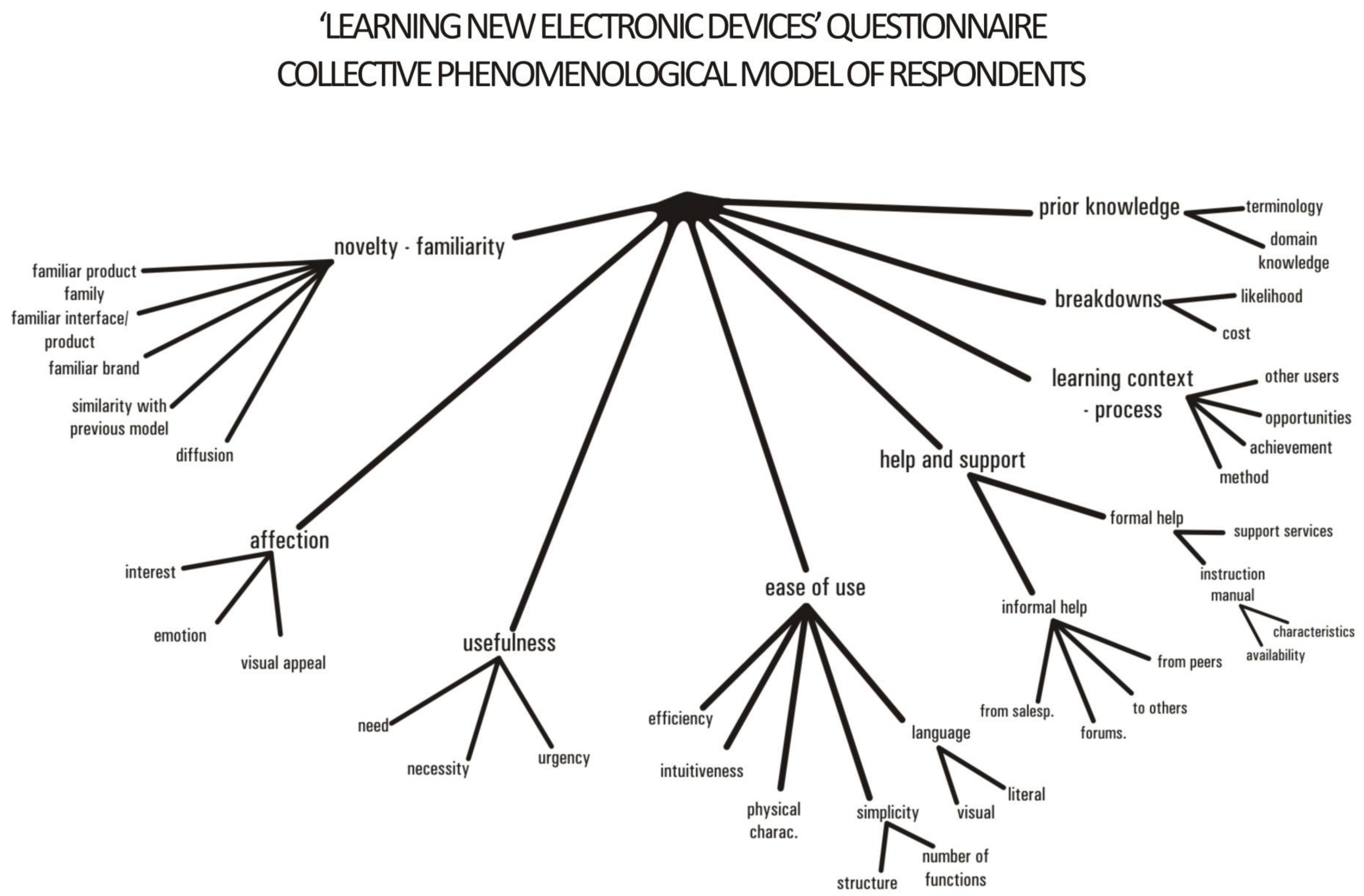


Figure 5.9 Phenomenological model after LEDQ

Together with the phenomenological model, it was observed that some of the expressions were related to "attempting to learn" and some were "capability to learn". Out of this differentiation a process model can also be derived. Detailed discussions about both models will be held in Chapter 6.

From the perspective of measurement, the distinction between 'not to attempt to learn' and 'attempts resulting in unsuccessful trials' is critical and worth consideration. If the data is examined in-depth, it may be suggested that problems witnessed by individuals with probably stronger self-efficacy beliefs are mostly related with 'not to attempt' because of certain disincentives. In order to contain such problems, the outcome of the decision process 'attempt?' should not be modeled as dichotomous, but should be modeled as to carry 'motivation' data as well. Then, it may be possible to suggest items such as 'I am confident that I can learn even an electronic device that I do not really need'. However, utmost care should be taken while working on items that primarily target *cluster I*, in order not to include 'will do' items instead of 'can do' items. Hence, items should be based on situations in which users decide to attempt a trial. Users' self-efficacy beliefs should be judged in presence of unfavorable situations and absence of favorable situations. Therefore, items should be focused on instances where learning process is broken or become too complex and demanding. In the table below there are some examples.

Table 5-2 Examples of item stems 1

Bir elektronik aleti...
"...takıldığımda yardım alabileceğim kimse olmasa da kolayca öğrenebileceğime inanıyorum." (Help and support)
"...üzerindeki ikonların (küçük semboller) ne anlama geldiğini anlayamasam da rahatlıkla öğrenebileceğime inanıyorum." (ease of use)
"...arkadaşlarımdan çok karışık bir alet olduğunu duymuş olsam bile kısa zamanda çok zorlanmadan öğrenebileceğimi düşünüyorum." (learning context and process)

Furthermore, it is apparent that the nodes suggested in the process model were not equally covered by the data collected. For example, although situations about the *feedback* after each trial were not mentioned by many respondents, items that target this loop may be generated.

Table 5-3 Examples of item stems 2

Bir elektronik aleti...
"...ilk denemelerim başarısız olsa da öğrenebileceğime inanıyorum."

The primary source for the generation of the item pool was the outcomes of this study. To put it more explicitly, 425 expressions derived with LEDQ were transformed into item stems after a selection procedure. Although in some cases expressions were directly worded as item stems, most of the times revisions in form and content were necessary. In the process of transformation, a set of criteria were applied in order to decide whether or not an expression will be utilized as an item stem, and whether or not a selected expression should be revised. These criteria were selected among several guidelines about item development for general purposes[24] and for self-efficacy scales specifically[25]. As previously explained both phenomenological and process models suggested after LEDQ were reflected in these guidelines.

[24] See Report II for a detailed discussion
[25] Bandura, 2006 and Bong, 2006

FORM

- Use of colloquialism or jargon should be avoided;
- Items should be clear, short, and simple;
- Items should ask only one situation to be evaluated at a time. Double-barreled items should be avoided;
- Double negatives should be avoided;
- Items with weak or very strong statements should be eliminated;

CONTENT

- Items should not force respondents to speculate on situations that they did not experience;
- Items should not ask for judgments based on experiencing a specific type of device;
- Items that denote situations which may enhance or hinder the learning process depending on respondents' personal characteristics should be eliminated[26];
- Items that suggest hard-to-generalize associations between situations and success in learning[27];
- Items that portray situations that affect whether the user will attempt to learn or not should be avoided[28].
- Items that target other kinds of self beliefs or inter-personal comparisons should be eliminated;
- Items that do not define a concrete situation should be eliminated;

---

[26] For example situations when the user needs to learn the device in a short time may either enhance the learning process, or may have a negative effect.

[27] For example items that include arguments about the appearence of the device were eliminated.

[28] Self-efficacy scales should contain 'can do' items instead of 'will do' items. See Report III for a detailed discussion

- Items should be context specific in order to avoid forcing respondents to base their judgments on abstract situations.

Some items with redundant wordings were kept so that these may be empirically evaluated in item tryout and major data collection. Some forms of colloquialisms were tolerated for the sake of avoiding the use of technical terms.

Besides these, expressions that are not related with the task of learning a new device and those that may not be associated with GISE were also discarded. The number of respondents that included the expression in their answers (frequency) was used as a reference. However, the decisions based on frequency values were not carried out in a strictly quantitative fashion. It was treated as an auxiliary criterion, especially in cases where an objective basis for making a decision was not present. Expressions with high frequency values were examined carefully even if they violate certain other criteria so that respondents' perceptions may be well represented, if criteria could be met by alternative wordings or slight modifications in the content. Expressions with low frequency (1) that are hard to accommodate within the collective phenomenological model were also scrutinized for relevance. Most of the time, such expressions were discarded for the sake of content validity.

5.5.3. Phenomenological model

It should be noted that especially collective phenomenological model[29] suggested does not necessarily reflect how respondents group situations that influence learning process positively and negatively. The category titles seldom reflect exact

[29] See *Report III* p. 12.

terms used by respondents and suggested to match common concepts in usability and related literature. Therefore, aim of the model is neither proposing a theoretical basis for GISE (General Interaction Self-efficacy) nor uncovering its inner structure. If the items grouped under each category are examined it is apparent that although some categories are homogeneous and have a distinct character, categories *learning context and process* and *prior knowledge* are quite heterogeneous. Although it was possible to subdivide these into smaller categories, numbers of items in these categories were not sufficient to prevent atomization. The heterogeneity was noted to be considered in following steps, so that diversity of content is conserved as much as possible.

At this stage, the primary utility of this phenomenological model was just to group similar items together, and to monitor the distribution of items which sample distinct content areas.

5.5.4. Wording

The wording strategy adopted was to simplify sentences and expressions as much as possible, without hindering the initial meaning. Furthermore so-called item hardness was tried to be adjusted with the use of proper wording. In doing so, the aim was to adjust statements in order to ensure that items are not rated with minimum or maximum scores by all of the respondents. Expressions were transformed so that each item stem was made up of a sentence depicting a negative situation, which is a frequently employed strategy in self-efficacy scales (see Bandura, 2006; Bong, 2006) Since respondents' self-efficacy beliefs regarding learning a new device in challenging conditions was to be measured, items were structured to convey meaning in the following patterns:

“Even if x is not present”,

“Even if x is present…”

Therefore, items were based on instances where positive factors are absent or negative ones are present. The following examples illustrate how expressions compiled in LEDQ were converted into item stems:

“Diğer aletlerden bildiğim kullanım mantığını uygulayabiliyorsam” > “Diğer aletlerden bildiğim kullanım şeklini uygulayamıyorsam”

“Çok kullanılan fonksiyonlar kolay bulunuyorsa” > “Çok kullanılan özellikleri kolay bulunuyorsa”

“Ürünün üstünde anlaşılmayan günlük hayatta kullanılmayan sözcükler varsa” > “Üstünde anlaşılmayan sözcükler varsa”

For the development of items of non-LEDQ origin, well established heuristics devised by Jacob Nielsen (Nielsen, 1994)[30] was utilized. Each guideline was critically evaluated for item generation potential. Most of the items generated this way, included concrete situations depicting undesirable interface characteristics. Expressions that contain such detailed descriptions about characteristics of interfaces were not observed in stems gathered in LEDQ.

“Hata uyarıları anlaşılmazsa.”

“Alet yaptıklarımı iptal etme şansı vermiyorsa.”

“Kullanım sırasında bir çok şeyi aklımda tutmam gerekiyorsa.”

[30] For an online copy and information about the updated list of heuristics see www.useit.com/papers/heuristic/heuristic_list.html

As a result, 242 items were generated to be evaluated by the experts. In the diagram below, content distribution before and after item generation is shown.

Table 5-4 Item distribution

Categories	Frequency in LEDQ (N*=425)	Frequency in item pool (N=242)	Δf[‡]
Novelty and familiarity	0.10	0.11	-0.01
Affection	0.08	0.08	0.00
Usefulness	0.08	0.10	+0.02
Ease of use	0.32	0.26	-0.06
Help and support	0.28	0.21	-0.07
Learning context and process	0.08	0.05	-0.03
Errors and breakdowns†	0.04	0.03	-0.01
Prior knowledge	0.04	0.03	-0.01
of non-LEDQ origin	-	0.14	-

* Total number of expressions / items

† Category was previously called 'breakdowns'

‡The difference between frequency values of expressions in LEDQ and item pool

With the introduction of items that are of non-LEDQ origin the weight of two major categories, namely ease of use and help and support were reduced by 13%. However, the category ranking according to frequencies is not drastically affected.

5.6. Expert review

The last item reduction done before empirical studies was done in accordance with evaluations made by a group of experts. Experts were also encouraged to suggest items, change or comment on the existing ones, which would broaden the content covered by item pool.

5.6.1. Methodology

242 items generated were submitted to 5 raters to be evaluated with regards to form and content. The following criteria were considered while choosing experts:

- Should be experienced in user research, specifically in the area of consumer products;
- Should be knowledgeable in concepts related to usability and interface design;
- Should be familiar to problems that user witness with digital interfaces;
- Should be experienced in usability testing;
- Should be experienced in preaparing and administration of questionnaires or similar paper-based data collection techniques

After the team of experts was assembled a document with following information was submitted together with the items to be evaluated:

- Rationale behind the main research;
- A short operational statement about the expected function of scale that will be developed;
- Detailed definitions about each keyword used in the operational definition;
- A brief description about the concept of 'self-efficacy';
- A brief description about the targeted construct 'General Interaction Self-Efficacy'
- Aim of expert review, how the results will be utilized
- Criteria of evaluation regarding the quality of wording (form);
- Criteria of evaluation regarding the validity of content (content);
- Technical notes about how scores and comments should be provided.

A sample of this document is provided in the Appendice C, D. After one of the raters asked for a detailed explanation about strategy to be adopted for scoring items, an e-mail was sent to all raters for further explanations. In this e-mail, experts are asked to reflect their own opinions in their 'content' scores and to evaluate each item on its own, without comparing it with alternatives and without considering the number of similar items. Furthermore an example about how the items will be presented to respondents was provided. Later on, some of the raters asked for more help about evaluation strategy. No extra expert training or applied instructions were given.

Raters were expected to evaluate each item with a 10-point scale ranging between 1 and 9. Response format enabled experts to submit 'neutral' scores (5).

It took approximately 4 to 8 weeks for experts to complete and return evaluation forms.

5.6.2. Results

Results of the expert review were provided in Appendix E.

Inter-rater reliability

Reliability among the scores provided by experts was calculated by correlating each rater's scores with the group average (Uebersax, 2000). Although correlation coefficients were inflated since each rater's score is reflected in both variables (rater's score, group average), reliability was quite low (r=0.54, r=0.55 for 'form' and 'content' scores respectively). If reliability was calculated in a conventional fashion so that scores of each rater is compared with other raters individually, coefficients were very low as expected.

Table 5-5 Inter-rater reliability

Form

	Rater A	Rater B	Rater C	Rater D	Rater E	Average
Rater A		0.08	0.14	-0.00	0.15	0.09
Rater B			0.15	0.14	0.15	0.13
Rater C				0.12	0.21	0.15
Rater D					0.12	0.09
Rater E						0.16
						0.12

Content

	Rater A	Rater B	Rater C	Rater D	Rater E	
Rater A		0.32	0.16	-0.07	0.17	0.14
Rater B			0.17	0.08	0.15	0.18
Rater C				0.11	0.28	0.18
Rater D					0.04	0.04
Rater E						0.16
						0.14

The fact that inter-rater reliability was low can be explained by the subjective nature of item evaluation, especially with regards to wording and differences in interpreting the construct GISE. Intra-rater correlation—i.e. correlation coefficients between form and content scores given by an individual rater—were quite high, ranging from 0.54 to 0.82, with an average of 0.63. The reason for such high values may be the fact that experts actually evaluated item quality as a whole, and then adjusted their scores considering form and content.

With these results, it was decided that item elimination should not be carried out totally based on average scores yielded by each item. The procedure will be discussed later.

Score distribution

Score distributions of individual experts are given below.

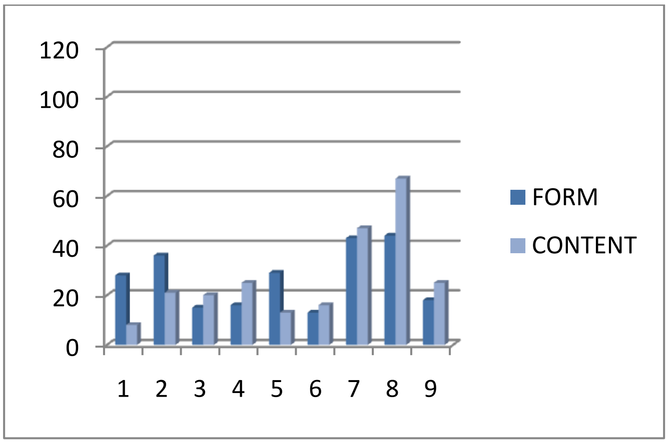


Figure 5-10 Score distributions of Rater A

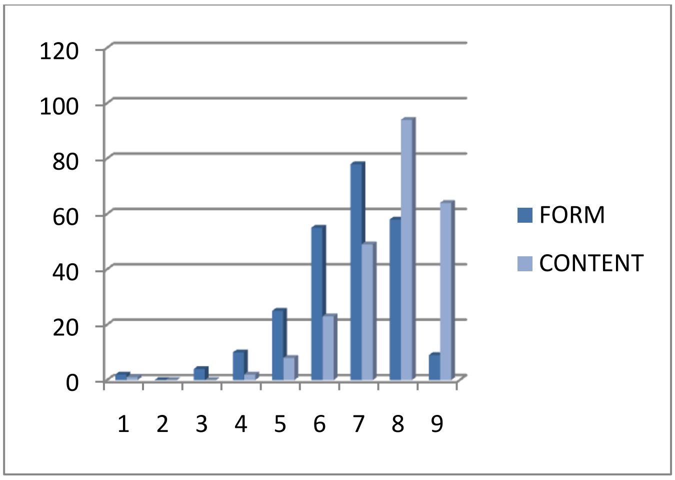


Figure 5-11 Score distributions of Rater B

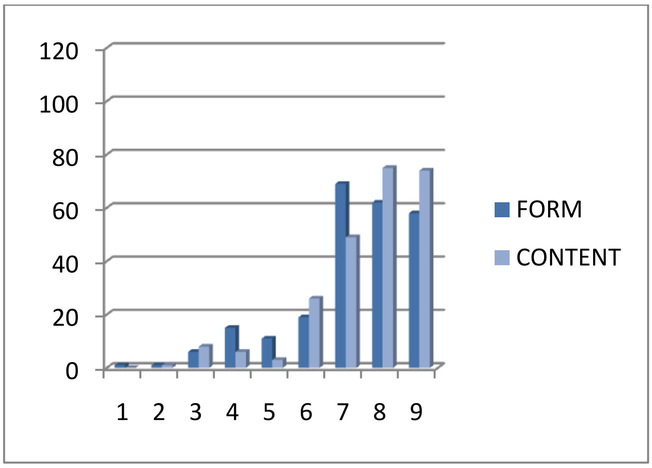


Figure 5-12 Score distributions of Rater C

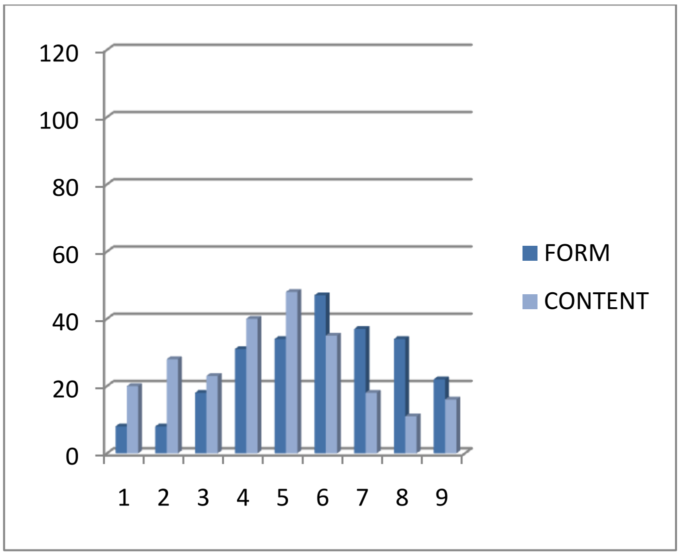


Figure 5-13 Score distributions of Rater D

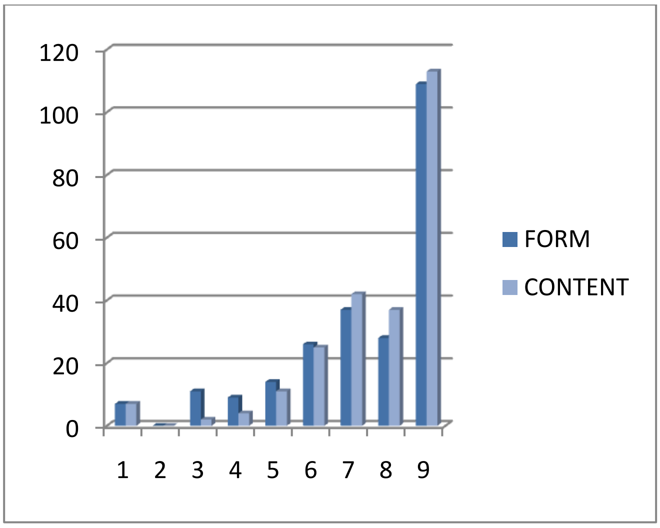


Figure 5-14 Score distributions of Rater E

Almost none of the distributions, except Rater D, were normal. Distributions for the raters B, C, and E were positively skewed with average scores quite higher than the expected midpoint.

Table 5-6 Mean, median and standard deviation values of scores submitted by raters

	Rater A		Rater B		Rater C		Rater D		Rater E	
	Form	Content	Form	Content	Form	Content	Form	Content	Form	Content
Mean	5.15	6.04	6.64	7.71	7.24	7.56	5.79	4.66	7.33	7.64
Median	5	7	7	8	7.00	8.00	6.00	5.00	8.00	8.00
St. Dev.	2.67	2.38	1.36	1.19	1.62	1.50	2.06	2.19	2.07	1.80

Average values across raters are 6.43 and 6.72 for 'form' and 'content' scores respectively. Together with common distribution characteristics; high average scores and low standard deviations made it necessary to determine some criteria to lead the item reduction process.

5.6.3. Item reduction criteria

Due to high average scores, low inter-rater reliability and relatively high intra-rater correlations, it was decided that form and content scores should be averaged and items to be eliminated should be somehow based on this composite score. Given the distribution characteristics, threshold was set to 6.50 instead of 5. However, items that yielded lower composite scores were also kept for further evaluation and both scores across raters and individual 'form' / 'content' scores were taken into consideration. The following points summarize the criteria that are utilized to systematically carry out reduction process.

Items with the following characteristics had the priority to be selected as a scale item:

- Items that yield a score of 6.50[31] or above;
- Items that yield a score below 6.50 in the presence of a single outlier[32];
- Items that have a low 'form' score, but a high 'content' score[33].
- Items that are derived from expressions observed with high frequencies in LEDQ;
- Items that play an important role in representing a sub-category[34];
- Items that fulfill item generation guidelines previously utilized.

[31] The composite value obtained after the 'form' and 'content' scores were averaged.

[32] Since inter-rater reliability is low there are many item scores where the average is quite high despite a single score below 3 (eg. 8-9-8-7-1). These items were also given priority in the selection process.

[33] Items that have a low 'content' score were not taken into consideration even they had an outstanding 'form'.

[34] Such items were tried to be improved by alternative wordings and reformulations.

Together with these, the item distribution characteristics summarized in were considered during item reduction, so that an imbalance among sub-categories is not created. This was done by determining quotas for each sub-category. However, theses quotas were not treated as strict limits, but as a framework to lead the elimination process.

5.6.4. Item reduction and the reduced item set

There were some defective items in the initial pool. These defects prevented consistent evaluation. Two of the item stems (13, 61) included positive expressions instead of negative ones. Although some raters submitted a score after correcting the items, 2 of the raters did not score item 13. Scores submitted to item 61 were complete. One item stem (210) included a double-negative statement.

113 and 116 were redundant items with exactly the same wordings. Therefore, item 116 was eliminated.

There were minor spelling mistakes but these did not hinder the meaning conveyed.

After the removal of defective items, item reduction process was carried out in line with the criteria listed above. The number of items was reduced from 242 to 104.

5.7. Major data collection

5.7.1. Materials and Method

Main Sampling Strategy

Required sample size for item try out and major data collection was determined as 50 and 450 previously. In order to ensure that the scale is administered to an unbiased sample, the sampling strategy was shaped in accordance with 3 points listed below:

- Sample should be composed of approximately 50% males and 50% females, reflecting the ratio in population[35].
- Age groups between 18 and 54[36] should equally be represented in the sample. Distribution should reflect real weights of the age groups in population.
- Every geographical region should be represented in the sample[37].

In accordance with these criteria sample population was defined as follows:

250 female and 250 male adults, resident in the districts of Çankaya, Yenimahalle, Mamak, Keçiören; between ages of 18 to 54…

---

[35] Although aim is not hypothesis testing with regards to the effects of gender, a severe imbalance should be avoided so that a possible source of a systematic error is eliminated.

[36] Age group partitioning employed by TÜİK is 18-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54. Therefore, 54 is set as the upper age limit.

[37] Sampling from a diversity of socioeconomic groups is tried to be attained by administering the scale in different districts of Ankara.

In order to determine the weight of age groups within sample population, data from TÜİK (Türkiye İstatistik Kurumu) was analyzed and the distribution was done to replicate the exact weights of the age and gender groups in Ankara population).The following table summarizes the distribution of age groups in Ankara (ADNKS, 2008) and how this structure is preserved in the sample population.

Table 5-7 Population and sample distribution to age groups

Age groups	Population	Males	Females	Ratio of age groups in population	Ratio of males in each age group	Ratio of females in each age group	Number of samples allocated to each age group	Males in sample population	Females in sample population	Total
18-24	511,803	268,871	242,932	0.27	0.53	0.47	134.3	71	64	134
25-29	308,493	153,919	154,574	0.16	0.50	0.50	80.9	40	41	81
30-34	270,499	133,383	137,116	0.14	0.49	0.51	71.0	35	36	71
35-39	268,515	132,858	135,657	0.14	0.49	0.51	70.4	35	36	70
40-44	225,234	112,881	112,353	0.12	0.50	0.50	59.1	30	29	59

Table 5.9 cont'd										
45-49	181,609	91,220	90,389	0.10	0.50	0.50	47.6	24	24	48
50-54	139,903	69,674	70,229	0.07	0.50	0.50	36.7	18	18	37
Total[38]	1,906,056	962,806	943,250	85.15	0.51	0.49	500.00	253	247	500

Sampling within districts

A strict sampling procedure such as determining the exact residences in which the scale will be administered was not employed. In order to make sure that certain sub-regions were not systematically visited more, streets were chosen randomly among all the streets that lie within the borders of the districts. Administrators were instructed for maintaining an unbiased approach in 'selecting' buildings to seek volunteers for participation. These instructions will be further discussed together with other instructions provided to administrators.

[38] Note that there are 554 males and 450 females in Ankara population with missing age data.

Administration

Scales were to be self-administered by respondents after a brief explanation of the task by the administrators. Study was carried out in residences, with only one resident at each residence. In order to ensure that required gender distribution is not very hard to attain, data collection in both item try out and final phase was carried out at weekends. Administrators first introduced themselves; explained the study, and how items should be scored using the rating scale. A short exercise was provided in order to familiarize respondents with rating items. Then, an informed consent was obtained from each respondent declaring that their participation is voluntary (see Appendix G). All the respondents were made sure that they can quit filling out the scale whenever they feel stressed either physically, or emotionally. Administrators left the respondent for approximately 30 minutes to 2 hours and returned back to pick up the scale. If the form was not completed administrators asked respondents to complete the form if they did not left it blank intentionally. In case where respondent refused to complete the form it was recorded as missing data and replaced with another administration.

Official permissions

Prior to data collection across 4 districts in Ankara, all the necessary permissions were requested from the following institutions:

- Middle East Technical University Human Subjects Ethics Committee;
- Governorship of Ankara;
- Ankara Department of Police.

Team of administrators

The team of administrators was assembled from a group of undergraduate and graduate students, studying in sociology in METU and Ankara University. Team consisted of four members who have a substantial amount of experience in administrating questionnaires and interviews in field studies.

Before the item try out, the team was subjected to a short training programme that consisted of 3 sessions. First two sessions lasted approximately 2 hours and the last session was a brief 30-minute meeting. In the first session, after discussing the team's previous experiences in field studies, a brief introduction about the area of research was presented. This was followed by a short presentation about the main research questions, the rationale behind the method to be employed, and how results will be utilized. After the session, hand outs that summarize the topics discussed were supplied. In the second session, administrators were introduced with the sampling strategy and the geographical regions where the study will be conducted. Furthermore, administrators were warned not to systematically choose a particular type of building (e.g. blocks, squatter's houses, etc.), exclude shops and any other kinds of work places in order to look for participants. Finally, administrators were instructed about the scale form, how should respondents be informed and problems that will possibly be experienced in the field. Before, the team was dismissed each district was assigned to a group of administrators. In the third session, an envelope that consisted of photocopies of legal permissions, scale forms, instructions, consent forms, district maps, and forms to record addressed visited was handed out to each administrator. After a final overview of the technique to be employed in the field, the team was dismissed.

At later stages of data collection, short informal meetings were held to discuss the problems experienced and strategic decisions to overcome these.

Scale form

104 items retained after expert review phase were included in this preliminary scale (see Appendix H). Further item reduction was expected after the initial item try out. Scale was composed of four parts:

- Questions that target demographics information (age, gender, level of education)
- Short instructions about GISE scale
- GISE scale items
- Checklist of electronic devices used by respondents[39].

A 0 – 10, 11 point scale was employed considering that respondents with low literacy may feel comfortable with submitting in the interval used in grading in formal education until 1990's.

The following rating scheme was employed with verbal anchors at both ends.

[39] Although scale development is the primary aim, additional information on the devices used by respondents were also collected so that an initial exploration about validity was done. In such a study, a moderate positive correlation between GISE score and the types of electronic devices used may indicate that the basic proposition "as users interact with more interfaces their GIE and therefore GISE increases" is valid.

Puanlama

0	1	2	3	4	5	6	7	8	9	10
Aleti öğrenebileceğime kesinlikle **güven<u>miyorum</u>**										Aleti öğrenebileceğime kesinlikle **<u>güveniyorum</u>**

Instead of putting a check to corresponding boxes, respondents were asked to write down scores in order to avoid careless and random responses to some extent.

1	Daha önce aynı işe yarayan bir aleti kullanmadıysam	Puan(0-10): _____

Since the scale form contained 104 items, it was suggested that possibility of careless responses would increase as respondent advances through the form. In order not to introduce a systematic error with regards to item orders, item set was partitioned into 5 sub modules (shown as A, B, C, D, E in Figure 5-15). 5 alternative forms (labeled as Form 1, Form 2, Form 3, Form 4, Form 5 in Figure 5-15) were prepared so that none of the modules were disadvantaged in terms of its order within the scale form.

Form 1	Form 2	Form 3	Form 4	Form 5
A	E	D	C	B
B	A	E	D	C
C	B	A	E	D
D	C	B	A	E
E	D	C	B	A

Figure 5-15 Item shuffle groups utilized in this study

Criteria for data reduction in item tryout

Criteria for data reduction were set as follows:

- Descriptive statistics in order to identify items with improper item difficulties[40] and unexpected variances[41];
- Items that are left blank frequently;
- Items that do not correlate with the rest of the items in the scale (i.e. items with low item-remainder coefficients).

[40] Item difficulty is used as a term to define sample mean of the scores yielded in a particular item. If the distribution is skewed to either hand, item is said to have low item difficulty (i.e. below expected mean—5 in this case) or high item difficulty (i.e. above expected mean).
[41] Variability of answers also regarded as a measure of good item design. Items with low variance are far from showing a discrimination power. For example, if all of the respondents rates an item with exactly the same score, this does not add anything to the measurement power of the scale. Therefore, deletion of such an item does not cause any loss of information.

Criterion 1 and 2 were set as auxiliary criteria for identifying potentially defective items. However, there are no conventional ways for an ultimate evaluation based on descriptive statistics and skipping behavior. Therefore, items that do not “pass” these two criteria were to be marked for further evaluation in later stages and especially against criterion 3. For criterion 3, as the main rule against which the item reduction was to be performed, a minimum acceptable value of 0.40 was set (Spector, 1992).

Hypotheses regarding independent and dependent variables

A preliminary analysis to explore relations between independent and dependent variables was done. In this regard, the following relationships were analyzed:

The number of electronic devices used by participants (NED) vs. total score calculated by the sum of scores yielded by all the items (Total Score)[42].

- Total score vs. age
- Age vs. NED

The expected type of relations by theory was a positive correlation between total score and NED, a negative correlation between total score and age, and finally a negative relationship between age and NED. In other words, it was hypothesized that individuals with higher total scores were expected to have a substantial experience with electronic devices. Besides this main expectation, it was hypothesized that younger individuals should have high total scores and should have a higher NED.

---

[42] Although the total scores are meant to reflect GISE-S score, at this stage, before the scale was developed by retaining superior items, it is early to name the total score as GISE

It should be noted that only the first relationship is a relationship between independent (NED) and dependent variable (total score). The other relationships were explored in order to explore further opportunities of providing proof of validity. Although the type of relationships in these two assumptions does not depend on previous theoretical discussions, face validity of both of these relationships are quite high.

5.7.2. Results of item tryout phase

Actual sample profile after data collection in item try out phase

Although not as strictly as it was in the major data collection phase, the sampling strategy previously discussed was tried to be maintained in item try out. In this respect, 65 scale forms were submitted to respondents and 62 forms were returned back to be analyzed. 10 of the cases were excluded due to following reasons:

- Missing demographical information;
- Pages systematically left blank, or forms with a considerable amount of unanswered items;
- Forms filled out in an unexpected way (e.g. respondent circles 0 or 10 in the rating label, ratings scores are totally illegible).

These misapplications were documented and reported to administrators in order to make sure that similar loss of data does not occur in the next phase.

After the elimination of defective forms ultimate sample size was 52.

The average age of the respondents was 33.2, with a minimum of 18 and a maximum of 55 (std. deviation = 11.2). 28 of the respondents were females and 24 of them were males. The geographical distribution of the respondents was 12, 9, 11, and 20 individuals in the districts of Çankaya, Yenimahalle, Keçiören and Mamak respectively.

Descriptive statistics

Mean values of the 104 items ranged between 3.90 (Item 55) and 5.63 (Item 42). These values were within ±1/3 standard deviations of the mean[43]. However, item 42 and 55 were reserved for further evaluation phases since deviation from the mean was significantly high regarding the other deviation values.

Variances ranged between 7.14 (Item 28) and 12.76 (Item 100) without any abnormally high or low values for any of the items.

With these results, no item reduction based on descriptive statistics was done, but item 42 was highlighted as a potentially defective item.

[43] Note that during literature research about scale development, it was not possible to locate a convention about how to interpret deviations from the expected mean. Therefore, an arbitrary border of ±1/3 standard deviations from the mean was determined. Together with this, outliers were searched manually even among the values within ±1/3 standard deviations from the mean.

Item-remainder coefficients

Item-remainder coefficients for the 104 items ranged between a minimum of 0.48 (Item 67) and a maximum of 0.92 (Item 51). Table below shows the rankings of items with respect to item-remainder coefficients.

Table 5-8 Item-remainder coefficients for the 104 items included in item tryout phase

Rank	1	2	3	4	5	6	7	8
Item no.	51	92	90	102	96	80	104	86
Item-remainder c.	0,92	0,87	0,86	0,85	0,85	0,84	0,84	0,84
Rank	9	10	11	12	13	14	15	16
Item no.	57	98	89	84	14	72	97	52
Item-remainder c.	0,84	0,84	0,84	0,84	0,83	0,83	0,83	0,83
Rank	17	18	19	20	21	22	23	24
Item no.	50	83	30	95	9	101	103	93
Item-remainder c.	0,83	0,83	0,83	0,83	0,82	0,82	0,82	0,82
Rank	25	26	27	28	29	30	31	32
Item no.	31	82	70	85	71	59	77	48
Item-remainder c.	0,82	0,82	0,81	0,80	0,80	0,80	0,80	0,79
Rank	33	34	35	36	37	38	39	40
Item no.	56	37	79	47	74	7	38	45
Item-remainder c.	0,79	0,79	0,78	0,78	0,78	0,78	0,78	0,77

Table 5-8 cont'd								
Rank	41	42	43	44	45	46	47	48
Item no.	76	2	43	100	3	46	75	88
Item-remainder c	0,77	0,77	0,77	0,77	0,76	0,76	0,76	0,76
Rank	49	50	51	52	53	54	55	56
Item no.	27	69	23	99	36	34	58	60
Item-remainder c.	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75
Rank	57	58	59	60	61	62	63	64
Item no.	39	4	44	32	53	24	49	40
Item-remainder c.	0,75	0,74	0,74	0,74	0,73	0,73	0,72	0,72
Rank	65	66	67	68	69	70	71	72
Item no.	1	12	81	5	6	54	55	16
Item-remainder c.	0,72	0,72	0,71	0,71	0,71	0,71	0,71	0,70
Rank	73	74	75	76	77	78	79	80
Item no.	8	19	94	66	73	91	29	11
Item-remainder c.	0,70	0,70	0,70	0,70	0,70	0,69	0,69	0,69
Rank	81	82	83	84	85	86	87	88
Item no.	22	61	62	68	10	18	63	35
Item-remainder c.	0,69	0,69	0,68	0,68	0,68	0,68	0,68	0,67
Rank	89	90	91	92	93	94	95	96
Item no.	65	33	21	78	87*	26*	64*	13*
Item-remainder c.	0,67	0,66	0,65	0,65	0,64	0,64	0,64	0,63
Rank	97	98	99	100	101	102	103	104

Table 5-8 cont'd								
Item no.	15*	41*	28*	17*	20*	42*	25*	67*
Item-remainder c.	0,59	0,58	0,58	0,57	0,57	0,52	0,51	0,48

Before data collection, reduction strategy was decided to be based on eliminating items below a certain value. The cutoff value for identifying defective items was determined as 0.40 (Spector, 1992). However, as shown in Table 5-9, all the coefficients yielded in this phase was above 0.40. Given the fact that it was not possible to identify defective items by evaluating the results of descriptive statistics, it was decided that the cutoff value should be increased so that some less reliable items are reduced in this phase. Although increasing the cutoff value may be thought to increase the probability of deleting non-defective items, Spector (1992) states that an item reduction strategy may be either based on a pre-determined cutoff value, or on number of items to be retained after the reduction process. In other words, one may either inter-item reliability may be the primary criterion, or the number of items to be included in the final scale may dominate the reduction strategy. Therefore, it may be deduced that, item-remainder coefficient threshold may be increased safely to some extent. In accordance with these, first cutoff value was set to 0.70. With this new threshold, 21 items would be eliminated. However, a closer inspection of items to be deleted revealed that some of the pre-determined categories would not be sufficiently represented or totally get lost (e.g. usefulness category) in the major data collection phase, if 0.70 was determined as the cutoff point. Given the fact that it is not methodologically safe to drastically alter the structure based on a study conducted on a relatively small sample (N=52), cutoff value was set to 0.65.

With the establishment of this criterion in a post-hoc fashion, it was possible to delete 12 items, without any drastic change in the pre-determined structure discussed in *Report III* and *IV*. Within this group of items, item 42, previously reserved for further evaluation given its high deviation value, was also reduced. However, item 55 was kept since item-remainder coefficient for this item was sufficiently high (0.71). As a result, scale was refined and a scale with 92 items was arrived at to be further refined in the major data collection phase.

Reliability

Although it is early to calculate reliability at this stage, since it is not known whether the scale is unidimensional or multidimensional, Cronbach alpha[44] was computed as 0.992, which also reflects the high item-remainder coefficients (see Table 5.9). The fact there were many redundant items utilized at this phase explains why the Cronbach alpha is above 0.90.

Content sampling after item reduction

After the item reduction done in this step, content sampled by items were summarized in Table 5-9.

[44] Cronbach alpha is a measure of inter-item reliability, ranging from 0.00 – 0.99 A higher alpha level indicates that on average items reliably measure the same construct. In social sciences an alpha level above 0.80 is considered a strong indication of reliability(e.g. Netemeyer, Bearden & Sharma, 2003).

Table 5-9 Content sampling in successive steps

Sub-category	1*	2	3	4
1.1 - Novelty – familiarity > familiar product family	4	4	2	2
1.2 - Novelty – familiarity > familiar interface / product	17	13	5	5
1.3 - Novelty – familiarity > familiar brand	6	4	0	0
1.4 - Novelty – familiarity > similarity with previous model	7	3	1	1
1.5 - Novelty – familiarity > diffusion	8	7	2	2
2.1 - Affection > interest	8	7	2	2
2.2 - Affection > emotion	20	15	5	2
2.3 - Affection > visual appeal	5	0	0	0
3.1 - Usefulness > need	20	16	8	6
3.2 - Usefulness > necessity	8	6	2	0
3.3 - Usefulness > urgency	7	3	1	1
4 - Ease of use [general]	36	21	6	6
4.1 - Ease of use> efficiency	8	4	1	1
4.2 - Ease of use> intuitiveness	28	21	8	6
4.3 - Ease of use> physical characteristics	15	3	2	2
4.4 - Ease of use> simplicity >structure	24	13	3	3
4.5 - Ease of use> simplicity >number of functions	8	6	1	1
4.6 - Ease of use> language >literal	14	6	4	4
4.7 - Ease of use> language >visual	5	0	0	0
5.1 - Help and support > informal help > from salespeople	6	5	2	2

Table 5-9 cont'd				
5.2 - Help and support > informal help > user forums	1	0	0	0
5.3 - Help and support > informal help > to others	3	0	0	0
5.4 - Help and support > informal help > from peers	26	24	7	7
5.5 - Help and support > formal help > instruction manual >availability	9	6	2	1
5.6 - Help and support > formal help > instruction manual > characteristics	66	30	9	8
5.7 - Help and support > formal help > instruction manual >support services	8	3	1	1
6.1 - Learning context and process >method	12	10	2	2
6.2 - Learning context and process >achievement	5	4	3	3
6.3 - Learning context and process >opportunities	7	6	1	1
6.4 - Learning context and process >other users	9	6	1	1
7.1 - Breakdowns>cost	9	4	2	2
7.2 - Breakdowns>likelihood	6	3	1	1
8.1 - Prior knowledge>terminology	4	4	1	1
8.2 - Prior knowledge>domain knowledge	6	4	2	1
Non-LEDQ	-	33	17	17

* 1 – LEDQ, 2 – Expert review, 3 – Item try-out, 4 – Major data collection

With the reduction of 12 defective items, only subcategory "Usefulness > necessity" was totally eliminated from the item pool. However, all the main categories remained in the content structure. The scale utilized in major data collection phase after item reduction is provided in Appendix H.

5.7.3. Results of major data collection phase

In the major data collection phase, 476 forms were returned by administrators. Nevertheless, 33 of the forms were eliminated. Some of the forms were excluded because of the similar reasons previously discussed in accordance with item tryout phase. In addition to these reasons, forms that contain even a single missing response to an item were also eliminated in order to have a dataset appropriate for factor analysis.

Ultimately, actual sample size in this phase was 442. The average age of the respondents was 33.3, with a minimum of 18 and a maximum of 58 (std. deviation = 10.5). 225 of the respondents were females and 218 of them were males. The geographical distribution of the respondents was 117, 107, 105, and 114 individuals in the districts of Çankaya, Yenimahalle, Keçiören and Mamak respectively.

Item remainder coefficients

Similar to the results in the item tryout phase item-remainder coefficients were quite high (see Appendix J). Only a single item (Item 70) had a considerably low coefficient (0.45) and was marked as a potentially defective item. Responses for this item ("Yanımda zaten o aleti kullanmayı üstlenmiş biri varsa") were quite variable when compared to the other responses. A close inspection revealed that some of the respondents considered the instance as a positive factor while others considered it as a negative one. Therefore, not only the magnitude, but also the direction of the responses to this instance showed great variance lowering the item-remainder coefficient significantly. The rest of the coefficients were above 0.65.

5.7.4. Exploratory factor analysis

As suggested in many scale development procedures (e.g. Netemeyer, Bearden and Sharma, 2003, in order to reduce items and explore the factorial structure of the item set utilized an exploratory factor analysis was conducted. One of the major reasons to conduct such an analysis was to explore the dimensionality [45]of GISE.

For determining the number of factors that underlie a construct, Netemeyer, Bearden and Sharma (2003) suggests that three criteria after factor analysis may be employed:

- Scree plot[46];
- Kaiser-Guttman principle[47];
- Comprehensibility of factors

[45] See Report IV for a brief discussion on dimensionality.
[46] According to scree plot technique, when eigen values are plotted against factors if a sharp decrease defined as an “elbow” may be detected, it is safe to conclude that number of factors before the “elbow” may adequately explain the majority of variance.
[47] According to Kaiser-Guttman principle, the number of factors with eigenvalues higher than 1.0 should be included.

After factor analysis was conducted[48], the "elbow" observed in the scree plot indicated that only a single factor solution may be safely chosen, which means that scale may be regarded as a unidimensional one.

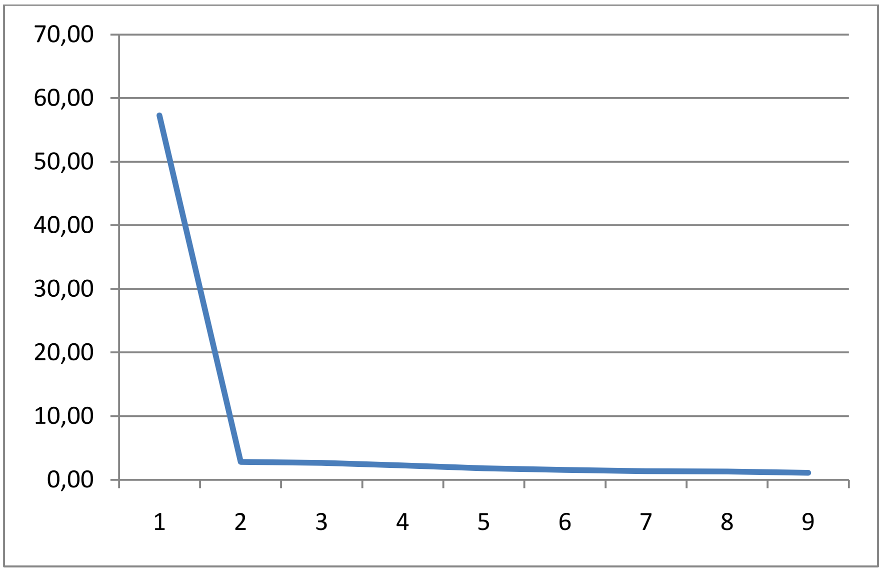


Figure 5-16 Scree plot after factor analysis

However, if Kaiser-Guttman principle was relied upon number of factors increased to 9. According to Netemeyer, Bearden and Sharma (2003) the ultimate decision should be made by considering comprehensibility of factors extracted.

[48] SPSS 17 was used for conducting exploratory factor analysis.

In order to check for theoretical comprehensibility of factors several factor solutions, starting from a 9-factor solution, were examined before deciding the number of factors to be extracted.

Only a single item ("70 - Yanımda zaten o aleti kullanmayı üstlenmiş biri varsa") was treated as an outlier since the item had considerably low item-remainder coefficients compared to the other items in the scale. The problem with the item was probably the possibility that some of the respondents treated the situation depicted in the item as a positive reinforcement while others treated it as a condition that affects the motivation to learn a device negatively.

In each factor solution the following set of item reduction criteria were utilized and the surviving items and factor structure was assessed with regards to their theoretical plausibility.

Factor analysis was done in accordance with the following main principles (Kleinbaum & Kupper, 1978):

- Simple structure and complexity reduction
- Independence among factors
- Conceptual meaningfulness and homogeneously sampled content

Operational criteria for reduction and assessment were as follows:

- Items that have loadings above 0.50 were considered significantly loaded by a factor[49].

---

[49] Since it is impossible to determine an absolute cutoff point was determined as 0.40. With this threshold it was not possible to eliminate items so that an-easy-to-administer number of items are retained. Depending on the 9-factor solution, the cutoff was increased until at least 5 items were retained in each factor group.

- Items that are loaded by more than one factor (above 0.40) were eliminated.
- Items that are theoretically irrelevant were eliminated even they comply with the other criteria.
- Factors should at least be loaded by 5 items in order to form a subscale.

9-Factor solution

A close inspection of the item groupings indicated that 9-factor solution is quite comprehensible (see Appendix K for factor loadings). When items included in these factors were evaluated it was evident that the preliminary phenomenological framework suggested was almost reflected in the factorial structure derived after the factor analysis.

However, after the item reduction was completed, factors 8 and 9 (breakdowns, learning context-process, and affection) were eliminated since there were no items significantly loaded by these factors.

8-Factor solution

In 8-factor solution, the factor structure resembles 7-factor solution after the elimination of factors 8 and 9. In this case 8th factor loads a single item (67), therefore 8-factor solution was also considered as inappropriate as far as a single item would not yield reliable results.

7-Factor solution

In this solution, factors 8 and 9 were totally eliminated. The remaining factors fit well with the theoretical categorization suggested after LEDQ.

6-Factor solution

In solutions where less than 7 factors were extracted many items were observed to significantly load more than one factor and both simple structure and theoretical comprehensibility was heavily compromised. Therefore, the assessment was terminated.

As a result, 7-factor solution was adopted. After the extraction of 7 factors and the employment of item reduction criteria defined above 66 items were retained in 7 subscales. However, for the sake of ease of administration, further elimination in order to have 5 items in each subscale was attained by removal of redundant items. Since all the items were above the cutoff values and complied with other criteria this last stage of reduction was not done based on quantitative means. In order to have a 7 x 5 structure items in each subscale were inspected with the help of item correlation matrix and redundant items were eliminated. The general strategy utilized was to reduce items without losing unique items that represent specific dimensions. Below is the final scale composed of 7 subscales.

Table 5-10 Subscale: Novelty

Familiarity – Novelty		Cronbach Alpha: 0.94
	Daha önce aynı işe yarayan bir aleti kullanmadıysam	
	Daha önce karşılaşmadığım bir aletse	
	Diğer aletlerden alıştığım kullanım şeklini uygulayamıyorsam	
	Daha önce alıştığım aletlerle arasında çok fark varsa	
	Kullanmaya alışık olmadığım teknolojiler içeriyorsa	

Table 5-11 Subscale: Motivation

Motivation (usefulness – affection)		Cronbach Alpha: 0.91
	Severek aldığım bir alet değilse	
	Kullanmaktan sıkılıyorsam	
	İşime yaramayacak özellikleri çoksa	
	Fazla ihtiyaç duymadığım bir aletse	
	Sıkça kullanacağım bir alet değilse	

Table 5-12 Subscale: Intuitiveness

Intuitiveness		Cronbach Alpha: 0.92
	Çok kullanılan özelliklerini bulmak kolay değilse	
	Hızlı bir şekilde istediğime ulaşamıyorsam	
	Sık sık kılavuza başvurmam gerekiyorsa	
	Mantık yürüterek çözebileceğim bir alet değilse	
	Temel özelliklerin nasıl kullanılacağı açık değilse	

Table 5-13 Subscale: Simplicity

Simplicity		Cronbach Alpha: 0.94
	Tuşlar birden fazla işe yarıyorsa	
	Çok fazla tuşu varsa	
	Menüsü çok karışıksa	
	Çok karmaşık özelliklere sahipse	
	Alet karmaşıksa	

5-14 Subscale: Informal help

Informal help		Cronbach Alpha: 0.96
	Satıcı nasıl kullanacağımı göstermezse	
	Bilen kişilere sorma şansım yoksa	
	Kullanımı gösterecek biri yoksa	
	Kullanabilen birini gözlemleme şansım yoksa	
	Takıldığım zaman yardım edecek kimse yoksa	

Table 5-15 Subscale: Formal help

Formal help		Cronbach Alpha: 0.95
	Kılavuzu yoksa	
	Kılavuz yeterince açıklayıcı değilse	
	Kılavuz anlaşılamıyorsa	
	Kullanım kılavuzunda günlük dilde kullanılmayan sözcükler bulunuyorsa	
	Teknik servisten telefonla yardım almak mümkün değilse	

Table 5-16 Subscale: Design

Specific design characteristics		Cronbach Alpha: 0.93
	Yaptıklarımın doğru mu yanlış mı olduğunu anlamakta zorlanıyorsam	
	Alet yaptıklarımı iptal etme şansı vermiyorsa	
	Ciddi sonuçlara yol açabilecek hata yapma ihtimali varsa	
	Ekranda önemli bilgiler net olarak verilmiyorsa	
	Hata uyarıları anlaşılmıyorsa	

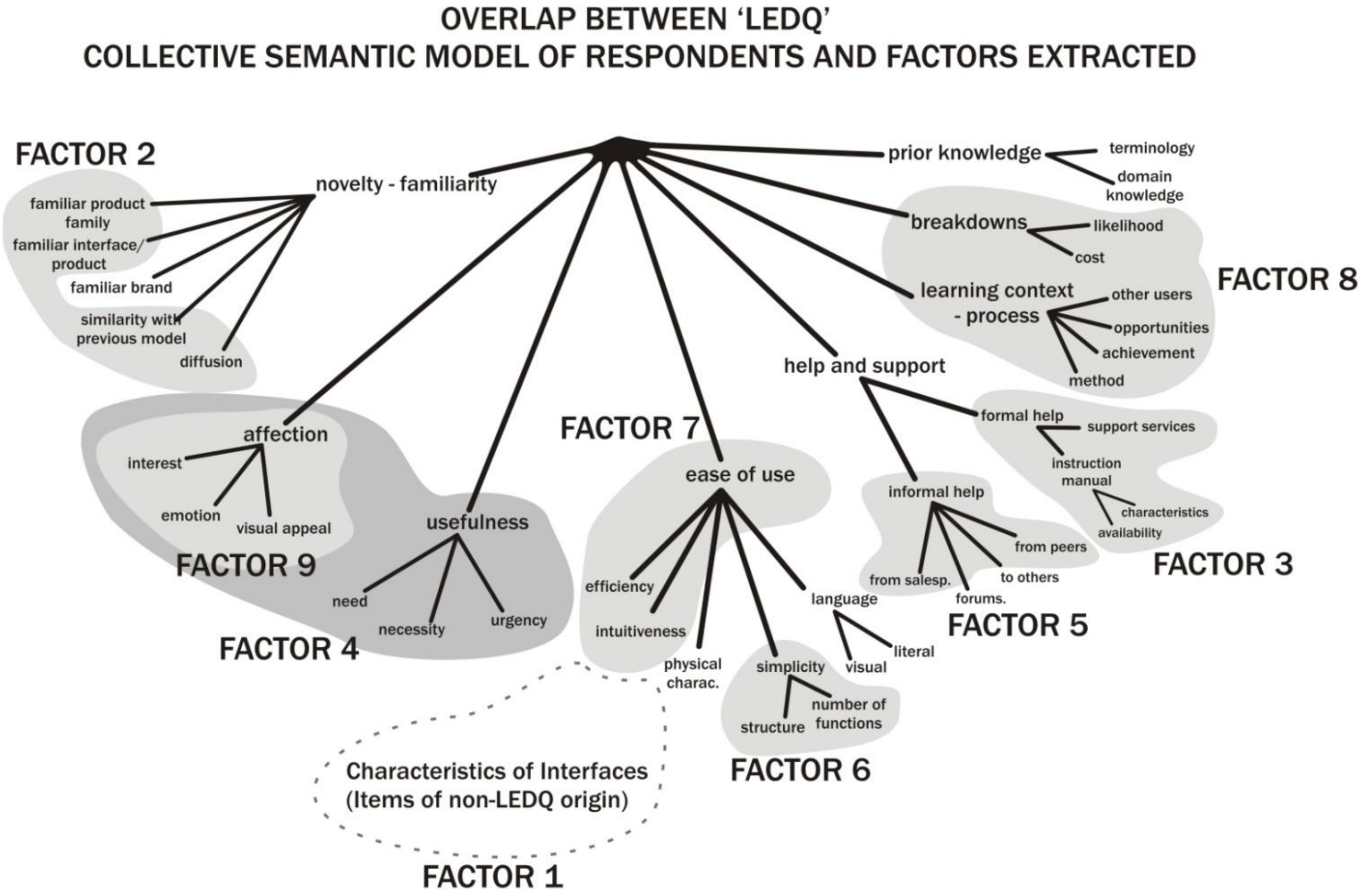


Figure 5-17 Overlap between phenomenological model and factors extracted

5.8. Validity studies

In order to provide evidence on the validity of GISE-S or in other words, to put forward what is measured by the scale is actually the construct defined as General Interaction Self Efficacy, some validity studies were conducted:

- One of these studies (Study 1) explored the relationship between GISE, NED, age, gender, district resided and education level.
- In order to provide an insight on predictive validity, two usability tests were conducted and effectiveness was compared with GISE scores (Study 2, Study 3).
- Finally, the structure of GISE was explored with SEM technique and alternative models were tested (Study 4).

5.8.1. Study 1: GISE and other variables

During major data collection, some additional data were gathered in order to conduct a validity analysis. These additional data consisted of age, gender, district resided, level of education and number of types of electronic devices experienced (NED).

Study 1A – GISE and Gender

In the first analysis the relationship between gender and GISE was studied. As it was discussed in the previous sections, gender is known to play a role in attitudes towards technology and computer use. Nevertheless, it is not too much to claim that gender causes differences in attitudes and it is observed that males usually have more positive attitudes towards technology and technology use. Although studying this phenomenon in detail is not within the aims of this study it was utilized in a known groups comparison fashion, in order to provide evidence regarding validity.

Hypothesis

H_1: Males do have higher levels of GISE if compared to females

Technique

One-way ANOVA was utilized in order to assess the relation between two variables.

There were 225 females and 218 males in the sample. The mean GISE for female respondents was 6.63 whereas mean GISE for male respondents was 7.30. This difference was found to be significant at 0.05 level (F=6.00; Sig. = 0,015) and null hypothesis was rejected.

Study 1B – GISE and Level of Education

In the second inferential study, the relationship between education level and GISE was examined. Although there is not much literature on this issue, it was expected that education level had an effect on GISE. However, it may be argued that this effect may be an indirect one, most probably moderated by NED.

Hypothesis

H_1: GISE will get higher as individuals' level of education increases.

Technique

One-way ANOVA was utilized in order to assess the relation between two variables. Level of education was represented with an ordinal variable with 6 levels. These levels were assigned as treatment groups:

1: no education, 2: primary school, 3: secondary school, 4: high school, 5: university, 6: graduate school.

There were no individuals in group 1 (no education). The descriptive statistics were provided in the table below:

Table 5-17 Sample population

Treatment group	N	Mean	S.D
1: No education	0	-	-
2: Primary school	28	3.93	1.49
3: Secondary school	44	5.46	2.57
4: High school	182	6.51	5.73
5: University	175	8.16	2.70
6: Graduate school	14	8.57	1.83

The differences between the means were shown to be significant at 0.01 level (F=24,96; Sig. = 0.00) and null hypothesis was rejected.

Study 1C – GISE and District Resided

In the third study exploring effects of readily observable variables on GISE, the effect of district resided was examined. Similar to education level, district resided was hypothesized to influence GISE indirectly. This effect may be suggested to be moderated by socioeconomic status, and therfore NED. In other words, it may be argued that as users have high socioeconomic statuses technology consumption rates increase and this may in turn increase GISE.

Hypothesis

H1: GISE will show difference across districts.

Technique

One-way ANOVA was utilized in order to assess the relation between two variables. District resided was represented with a nominal variable with 4 categories. These categories were assigned as treatment groups:

Table 5-18 Distribution across districts

Treatment group	N	Mean	S.D
1: Çankaya	117	7.82	2.98
2: Yenimahalle	107	6.83	2.60
3: Keçiören	105	7.42	3.00
4: Mamak	114	5.77	2.54

The differences between the means were shown to be significant at 0.01 level (F=11.67 ; Sig. = 0.00) and null hypothesis was rejected.

Compared to other findings that explore known groups comparison, difference between the means with regards to district resided is a controversial one. First of all, with only the district info, this finding is only meaningful on local basis. The differences between the districts on the basis of average income, education level and other socioeconomic indicators should be explored.

Study 1D – GISE, NED and Age

In the fourth analysis the relationship between age, NED and GISE was explored. As it was determined in the preliminary studies, GISE is positively correlated with NED and negatively correlated with age.

The Pearson's r between age and GISE was found to be -0.31, whereas r between GISE and NED was 0.46. As expected, there was also a negative correlation between age and NED (-0.35). In other words, respondents with high GISE were younger individuals who use more electronic devices.

In order to control the effect of age and isolate the effect of NED on GISE a partial correlation was run. Results indicate that when controlled for NED the correlation between GISE and age decreases to -0.17, therefore it is safe to claim that GISE is mainly affected by NED rather than age. When controlled for age, the correlation between GISE and NED was decreased to 0.40. Although there was a 0.06 point decrease, this value still indicates a high level of correlation.

Compared to other studies these results serve two purposes. As it is the case with other results, showing that GISE is negatively correlated with age gives opportunity for known groups comparison. Besides this, showing that GISE and NED are closely correlated and the effect of age considerably decreases when controlled for NED is an evidence for construct validity and a partial justification of triadic model suggested in this study. However, it should be noted that additional data is needed to verify these relations.

5.8.2. Study 2: GISE-S and Usability

As it was stated before, both the prototypical apparatus tests and GISES were developed in order to control individual differences based on individuals' expertise in interaction with digital products, in the case of usability tests. In line with this, definitions for both GIE and GISE are based on individuals' competencies in coping with "a novel interaction situation". Similar to the preliminary validity studies conducted for studying the relationship between performance in a usability test and apparatus test scores, a usability test was organized for exploring the predictive validity of GISES.

Hypothesis:

It was hypothesized that there should be a positive correlation between performance in a usability test and GISES scores.

Material and method

Selection of product to be tested in the usability test

Prior to selection of the test object, a set of criteria was determined to ensure that the product was appropriate regarding the aim of the study:

- The test object should be a consumer product.
- For ensuring versatility it was decided that the test object should be portable and should not require any sort of installation.
- For controlling prior experience so that “a novel interaction situation” is attained, the test object should not be a commonly experienced product.
- In order to minimize the effects of domain expertise, the object should belong to a widely used family of products.
- For maximizing “the novelty” of the interaction situation, interface of the test object should have uncommon characteristics.

In accordance with the criteria listed above, a Motorola cellular phone was selected within a set of 10 alternatives. Alternatives were as follows:

- Electrolux microwave oven;
- Panasonic dect phone;
- HTC Touch 2 pro PDA phone;
- Trimax DVD player
- SONY music set
- VESTEL television set with an OSD
- Packard Bell mp3 player
- Canon EOS 40D digital camera
- Canon HD video camera
- Motorola Cellular Phone

Tasks

12 scenarios were developed and 7 were selected to be included in the test. Selection of tasks was based on following criteria:

- Scenarios should not contain tasks that require specific knowledge that may render certain participants advantageous over others. In this regard settings that are specific to the product or tasks that necessitate domain specific knowledge were avoided.
- Tasks that require much time or activity were not included in order to limit what is experienced in each task. Tasks that require more than 1 minute were eliminated after expert efficiency values were determined[50].
- Scenarios that require a prerequisite task to be completed were not issued.

The following tasks were determined in line with the above criteria[51]:

Task 1. Participant was asked to find an entry from the phone book

Task 2. Participant was asked to send an SMS containing the message "Merhaba nasilsin?" to a person recorded as "ALICEP".

Task 3. Participant was asked to create a new contact in the phone book (Mehmet Kara: 0 555 220 20 20).

Task 4. Participant was asked to take a photo and find the associated file after returning to main menu.

Task 5. Participant was asked to assign a photo to an entry in the phone book.

Task 6. Participant was asked to display the remaining credit

Task 7. Participant was asked to setup time and date to 13:30 – 15.05.2009.

[50] See Determination of Time-out Threshold Values
[51] The contents of scenario cards used in the tests were provided in the Appendix.

Determination of time-out threshold values

It is known that some individual differences are observed regarding when a participant quits a task or how an individual explores the interface while trying to attain the goals in a usability test. Some individuals may be inclined towards quitting a task after an unsuccessful attempt whereas some feel challenged and are motivated to keep trying until moderator somehow terminates the task. In this regard, determination of time-out thresholds based on empirical values was crucial in order to limit what was experienced by each participant after a task.

Values were determined by calculating the average time required to complete each task by two expert participants in three trials. Expert participants were given step-by-step instructions and completed each task three times and it was ensured that participants were fluent enough to be regarded as expert participants.

Procedure

Steps of the procedure followed in the test are listed below:

Screening of the potential participants: Screening was made in order to ensure that participant was between 25 and 35, was at least a university graduate, uses PCs on a daily basis, and has no experience with the cellular phone to be tested.

Administration of GISE-S: Scales (see Appendix M) were self-administered without any verbal instructions. Written instructions and an example were provided with the scale form. It was ensured that all the participants administered GISE-S before the usability test.

Instruction about the usability test: An explanation about how the test will be conducted was provided in order to ensure that participants will not experience any problems due to the way test is conducted. Participants were especially informed about the "time-outs".

Administration of the usability test: Participants were not recorded during the test. Simultaneous logging of the data was made by the facilitator. Only effectiveness and efficiency was measured during the test. Time was kept with a stop watch.

Contacts, messages and photos taken during each session were deleted and phone was reverted to the default time and date.

Sample population

In order to control the effect of age, education, computer literacy and gender, which are known to affect performance with a digital product, a quite narrow sampling scheme was adopted. The following points summarize the strategy followed during sampling:

- Participants should be between 25-35;
- Sample population should not be heterogeneous regarding level of education;
- Sample should not be biased regarding gender,
- Participants should have no prior experience with the specific product being tested;

- Participants should have a considerable level of computer literacy;
- Participants should be sustaining their work routines with PCs.

Operationalization of measures

Since the study aims to explore a correlation between usability test performance and GISE, two representative variables were defined.

Performance in a usability test was represented with effectiveness after 7 tasks. If a participant was able to complete a task by attaining the pre-set goals, effectiveness score for that task was recorded as 1. If a participant quits the task, exceeds the time-out values or thinks that the task was accomplished although it is not, effectiveness score was regarded as 0. Effectiveness for each task was operationalized as a dichotomous variable, that is, no means for determining partial effectiveness was suggested.

GISE was represented with the sum of the ratings after completing GISE-S. In order to conduct further analyses, sub-scale scores were also calculated.

Results of the study

The mean effectiveness yielded by participants after 7 tasks was 0.55, that is, roughly 50% of the tasks were not completed successfully. The lowest UP (compound effectiveness) was 1 out of 7 tasks (0.14), whereas the highest UP value attained was 6 out of 7 tasks (0.86). GISE-S scores ranged between 161 and 314, with a mean value of 233.83. As far as the highest possible score was 350, it may be regarded as a high value. However, since no normative data is present at the moment, such an interpretation may not be plausible.

Although the sample size is extremely small, the correlation between usability test performance (UP) and GISE-S scores was significant at 0.01 level (r = 0.93). As expected, negative correlations between Age - UP and Age – GISE-S were observed, however these were not significant.

Table 5-19. Results of the usability test and GISE-S

Task	U1	U2	U3	U4	U5	U6	U7	U8
Finding a phone no.	TO[52]	0:28	TO	TO	0:29	TO	0:22	Quit[53]
Sending an SMS	2:13	TO	TO	1:30	1:20	1:15	TO	3:00
Creating a new entry	1:33	0:30	1:37	0:27	0:43	1:08	1:07	TO
Taking a picture	TO	Quit	TO	2:30	TO	1:03	1:22	TO
Finding the picture	0:40	0:33	TO	0:50	0:34	0:31	TO	TO
Displaying remaining credits	TO	Quit	TO	TO	TO	0:19	TO	TO
Setting up date and time	0:40	TO	TO	TO	0:49	1:22	TO	2:00
UP* (Out of 7)	4	3	1	4	5	6	3	2
GISE-S score	212	187	161	261	268	314	223	195

[52] TO: Time out; Quit: User quited before success of timeout
[53] TO: Time out; Quit: User quited before success of timeout

Table 5-20 Correlations between variables

		Age	UP	GISES
Age	Pearson Correlation		-,420	-,481
	Sig. (2-tailed)		,300	,228
	N		8	8
UP	Pearson Correlation	-,420		,929**
	Sig. (2-tailed)	,300		,001
	N	8		8
GISES	Pearson Correlation	-,481	,929**	
	Sig. (2-tailed)	,228	,001	
	N	8	8	

**. Correlation is significant at the 0.01 level. Age: Age of participant, UP: Usability test performance, GISES: General Interaction Self Efficacy Scale Score

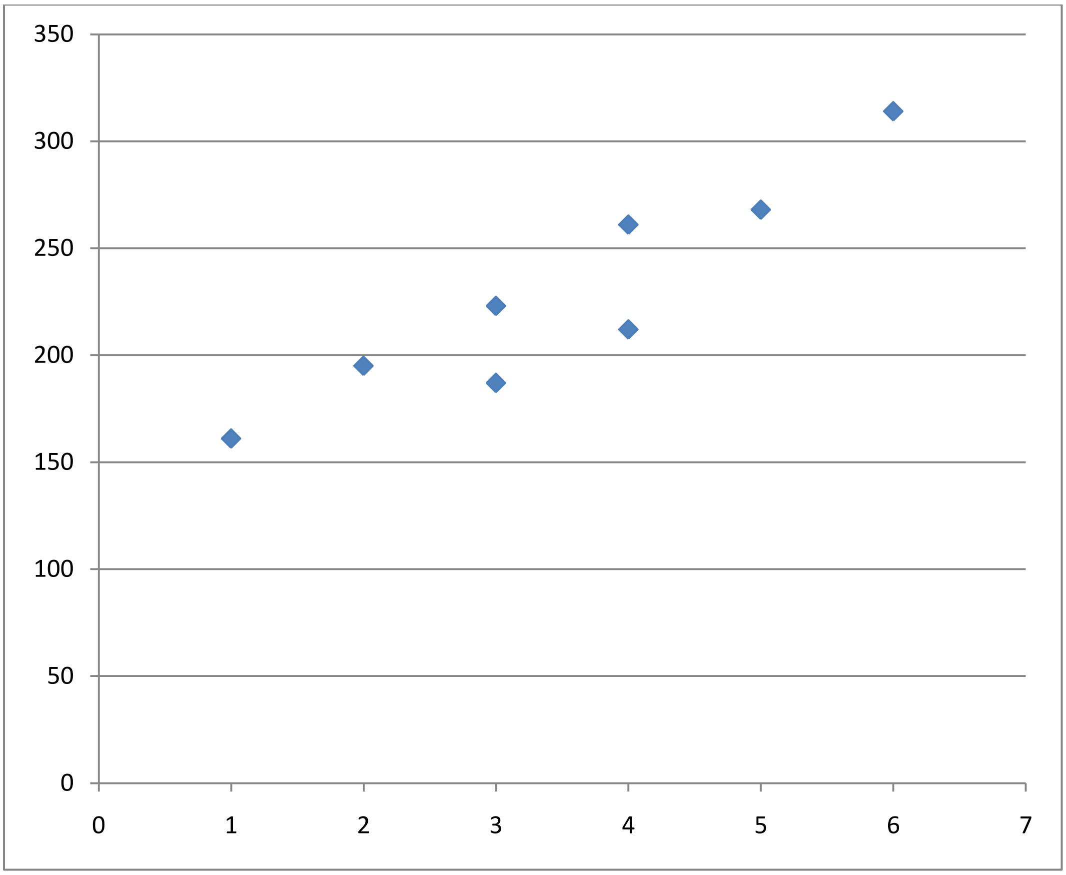


Figure 5-18 GISE-S vs. UP

Since interpretation of efficiency values are quite problematic any analysis on efficiency values was not done.

Table 5-21 Subscale scores and their correlations with UP

		UP
Novelty	Pearson Correlation	,678*
	Sig. (1-tailed)	,032
	N	8
Motivation	Pearson Correlation	,665*
	Sig. (1-tailed)	,036
	N	8
Intuitiveness	Pearson Correlation	,879**
	Sig. (1-tailed)	,002
	N	8
Simplicity	Pearson Correlation	,759*
	Sig. (1-tailed)	,014
	N	8
Infhelp	Pearson Correlation	,696*
	Sig. (1-tailed)	,028
	N	8
Formhelp	Pearson Correlation	,945**
	Sig. (1-tailed)	,000
	N	8
Spdesignch	Pearson Correlation	,914**
	Sig. (1-tailed)	,001
	N	8

* Correlation is significant at 0.05 level.
**Correlation is significant at 0.01 level

When the correlations of each subscale score to UP is considered, it is observed that all the correlations were significant. The lowest correlation was observed between UP and motivation. These findings should be systematically explored with further studies.

5.8.3. Study 3

Similar to the validity study "Study 2" GISES was administered in a real-life usability test to further explore the predictive validity of GISES.

Hypothesis

It was hypothesized that there should be a positive correlation between performance in a usability test and GISES scores.

Material and method

Although the usability test was a real-life one, the product tested complied with the criteria defined in the previous study. The test object was an IP (Internet Protocol) TV set-top box, used with a remote control and a TV set. In addition to the conventional TV features, system included VOD (video on demand). The interface was a full-screen GUI utilized by navigation controls and color-coded buttons[54].

[54] No additional information can be given about the interface due to Non Disclosure Agreements.

Tasks

8 scenarios were defined and included in the test. Selection of tasks was based on interests of the manufacturer and research design, so that no control over scenarios was possible.

The following tasks were administered during tests:

Task 1. Participant was asked to turn on the system.

Task 2. Participant was asked to switch to a channel.

Task 3. Participant was asked to find TV programme info for two channels using EPG (Electronic Programme Guide).

Task 4. Participant was asked to set a reminder for a TV programme using EPG, and then cancel it.

Task 5. Participant was asked to search a movie by name in the free VOD movie archive.

Task 6. Participant was asked to look for a movie by genre among movies to be rented.

Task 7. Participant was asked to find and watch a missed TV series.

Task 8. Participant was asked to form a favorites list and then zap among them.

Determination of time-out threshold values

In line with the first study, time-out thresholds were determined in this study as well.

Values were determined by calculating the average time required to complete each task by two expert participants in three trials. Expert participants were given step-by-step instructions and completed each task three times and it was ensured that participants were fluent enough to be regarded as expert participants.

Procedure

Steps of the procedure followed in the test are listed below:

Screening of the potential participants: Screening was done in order to have a participant profile consistent with manufacturer's target population. Therefore, no control was possible at this step.

Instruction about the usability test: An explanation about how the test will be conducted was provided in order to ensure that participants will not experience any problems due to the way test is conducted.

Administration of the usability test: Participants were recorded during the test. Simultaneous logging of the data was made by the facilitator. Effectiveness, efficiency was measured and problems were logged during the test. Measurements were refined after the test with observation software.

After each session, system was reset and reverted to the initial settings.

Because of the initial research design, participants had to fill in GISE-S after completing the test.

Sample population

Participants were between 25 and 35. The gender distribution was 50% and 7 of the participants were cable TV subscribers, whereas 5 of them were accustomed to digital platforms or satellite receivers.

Operationalization of measures

As it was in the previous study, since the study aims to explore a correlation between usability test performance and GISE, two representative variables were defined.

Performance in a usability test was represented with effectiveness after 8 tasks. If a participant was able to complete a task by attaining the pre-set goals, effectiveness score for that task was recorded as 1. If a participant quits the task, exceeds the time-out values or thinks that the task was accomplished although it is not, effectiveness score was regarded as 0. Effectiveness for each task was operationalized as a dichotomous variable, that is, no means for determining partial effectiveness was suggested.

GISE was represented with the sum of the ratings after completing GISE-S (see Appendix M). In order to conduct further analyses, sub-scale scores were also calculated

Results of the study

The mean effectiveness yielded by participants after 8 tasks was 0.62, that is, 62% of the tasks were not completed successfully.

Table 5-22 Results of the usability test and GISE-S

Task[55]	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12
UP	6.00	5.00	1.00	7.00	ND[56]	6.00	4.00	3.00	4.57[57]	6.00	4.00	8.00
Cont												
GISE-S score	166.00	162.00	125.00	261.00	ND	282.00	85.00	181.00	297.00	219.00	120.00	256.00

UP: Usability test performance, compound effectiveness scores

[55] Order of scenarios were shuffled and no Scenario number information was provided in order to comply with non-disclosure agreements.
[56] Data for this participant was eliminated since it was revealed that participant scored GISE-S items specifically for the product being tested.
[57] One of the scenarios could not be completed because of system breakdown.

The lowest UP (compound effectiveness) was 1 out of 8 tasks, whereas the highest UP value attained was 8 out of 8 tasks. GISE-S scores ranged between 85 and 297, with a mean value of 195.92.

Although the sample size is small, the correlation between usability test performance (UP) and GISE-S scores was significant at 0.05 level (r = 0.61).

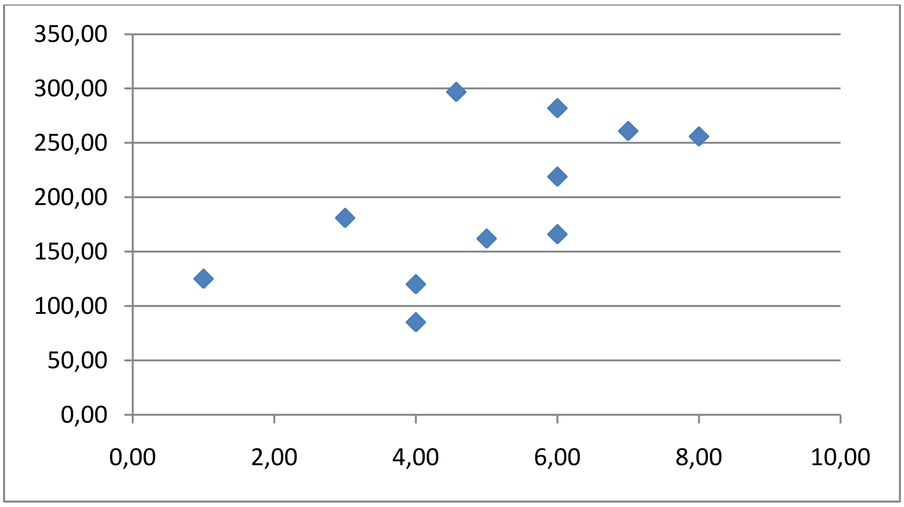


Figure 5-19 GISE-S vs. UP

As it was discussed in the Study II, since interpretation of efficiency values are quite problematic, analysis on efficiency values were not done.

Table 5-23 Subscale scores and their correlations with UP

		UP
Novelty	Pearson Correlation	,280
	Sig. (1-tailed)	,202
	N	11
Motivation	Pearson Correlation	,542*
	Sig. (1-tailed)	,042
	N	11
Intuitiveness	Pearson Correlation	,229
	Sig. (1-tailed)	,249
	N	11
Simplicity	Pearson Correlation	,516
	Sig. (1-tailed)	,052
	N	11
Infhelp	Pearson Correlation	,786**
	Sig. (1-tailed)	,002
	N	11
Formhelp	Pearson Correlation	,608*
	Sig. (1-tailed)	,024
	N	11
Spdesignch	Pearson Correlation	,662*
	Sig. (1-tailed)	,013
	N	11

* Correlation is significant at 0.05 level.

** Correlation is significant at 0.01 level.

When the correlation coefficients of each subscale score with UP are considered, it is observed that the significant correlations were attained by the subscales *motivation, informal help, formal help specific design characteristics.* The lowest correlation was observed between UP and intuitiveness.

In both studies presented above, GISE-S scores were correlated with usability test performance in the expected direction. It was shown that participants having high GISE-S scores performed well in usability tests and participants with low GISE-S scores were mostly poor performers. This relation was observed to be a very strong one in Study 2 (r = 0.95) whereas proportionality was weaker in Study 3 (r = 0.61). Despite this difference, r value yielded in Study 3 may also be regarded as a high value in the field of social sciences.

Besides the fact that both values were high enough to indicate a strong relationship and provide evidence for predictive validity, what may have caused this difference will be discussed in Chapter 6.

5.9. Study 4: Structure of GISE

Up to this point, GISE was handled within a measurement perspective, as an aggregate score to represent a user's self-efficacy beliefs. Therefore, in the validity studies, GISE was treated as a single variable and was correlated with corresponding variables. Although this treatment is plausible with regards to have a parsimonious, simple model; it was thought that exploring how sub constructs of GISE relate to each other may make it possible to gain insights about the phenomenon and the process of building GISE.

With the purposes of building a model that reveals the structure of GISE and how sub constructs are related to each other, Structural Equation Modeling (SEM) technique was employed.

According to Jöreskog & Sörborm (1993; also ctd. in Şimşek, 2007), SEM may be utilized with regards to three research strategies.

(1) A strategy for confirmatory purposes may be adopted by the researcher so that, a clear and well-defined model may be tested for confirmation.

(2) A second strategy is defined as alternative models strategy where a number of models are checked as to find out the best-fitting model.

(3) Model building may be a third strategy to find out best-fitting model and refine it in order to arrive at an ultimate model. With this strategy partial models may be developed and then nested in a main model.

The strategy adopted in this study is both a generative and an evaluative one. From generative perspective, results of the scale development process were tried to be explored in order to arrive at a deeper understanding of the construct defined as GISE. From the evaluative perspective, theoretical appropriateness or comprehensibility of the model developed would be helpful in providing evidence for the construct validity.

With these concerns in mind, a two-step modeling approach was adopted (Kline, 2005). Before testing alternative structural models and determining the best fitting model, measurement model was studied and refined.

5.9.1. Theoretical background in the model building process

Before testing the measurement model, seven factors extracted after exploratory factor analysis were evaluated and a structural model was specified. Latent constructs which cannot be theoretically related to other constructs were left undefined at this stage. In the following lines each latent construct was discussed regarding how they can be handled in the model building process.

NED

In line with the triadic model proposed in this study, number of electronic devices experienced by users (NED) was assigned as the only independent variable, consisted of a single observable variable. There is both theoretical and empirical evidence in order to safely state that there is a directional relationship between NED and GISE, where NED is independent and GISE is a dependent variable.

Formal Help

Among the factors extracted *formal help* was determined to be inappropriate to be included in the structural model, since it may be claimed that reading instruction manuals is a matter of personal style and most of the users do not refer to instructional material (e.g. Novick & Ward, 2006; Rettig, 1991) regardless of their level of expertise. Although it was utilized as a subscale within the measurement perspective, theoretically it is hard to specify the relation of this sub construct to other ones. In other words, although belief in ability to learn a new device without the presence of formal help may be regarded as a sign of high GISE for some users, act of referring to instruction manuals may not be related with GISE or a stage in the GISE development process. In order *formal help* to be included to the model, more theoretical and empirical findings are necessary.

Intuitiveness

Intuitiveness is a trait of interfaces that are easy to use and is valuable especially for novice users (Cooper & Reimann, 2003). Intuitiveness is a goal for good interface design where minimal knowledge or experience is assumed in the user's side, so that user may interact with the product almost instinctively. For example, it is suggested that walk up-and-use-products should be intuitive ensuring that no prior experience or training is necessary for first and one-time users (ISO 20282). Therefore, it may be stated that belief in ability to cope with non-intuitive interfaces may be regarded as the first step towards building self-efficacy beliefs. In other words, it may be suggested that users who believe that they are able to learn intuitive interfaces but not more complex ones may be in the preliminary stages of building GISE beliefs.

Complexity, Novelty, Design (Design characteristics)

By definition, belief in ability to cope with novel interaction situations, where individuals come across with complex products that may bear unfavorable design characteristics were suggested as sign of somewhat developed GISE. Compared to *intuitiveness*; *complexity, novelty* and *design characteristics* may be regarded as targeting the core of GISE. In other words, it is plausible to suggest that as individuals start to build GISE, they would most probably build beliefs regarding intuitive interfaces first but would experience problems with the ones that are novel, complex and composed of design characteristics that hinder ease of use.

Others (Informal help) and Motivation

Interpreting and specifying self-efficacy beliefs on informal help with regards to a level or stage of GISE seems to be problematic compared to others, although it is observed that experts mostly learn on their own and help others (Kiesler, Zdaniuk, Lundmark, & Kraut, 2000) and this is a form of strengthening social position (Ribak, 2001). It is argued that self-efficacy beliefs may flourish if the environment is not supportive (Compeau and Higgins, 1995; ctd. in Wu & Rocheleau, 2001) indicating that self-belief in coping with challenging situations is definitely an important aspect of GISE. However, whether this is a cause or effect cannot be safely assumed at the moment, even it seems plausible to argue that dependence to others in the process of learning an electronic device may be associated with individuals with low GISE or individuals that are new in GISE building process.

As it may be recalled, *motivation* was revealed as a composite factor that corresponds to situations where lack of usefulness and affection is present. Similar to depending on others for learning, belief in ability to learn new electronic devices even if they are not useful or emotionally attractive for a user may both be a cause or an effect. In other words, “ability” to learn a new electronic device even if it is not seen useful or emotionally satisfying may help one to build GISE quickly, or this belief may be a result of strong self-efficacy beliefs. The fact that high self-efficacy beliefs determine what an individual experiences and is a strong motivation in itself for dealing with corresponding activities probably indicate that *motivation* may mostly be an effect.

The core model

In the figure above, a core model to be explored and further specified with SEM technique was proposed. The core model specifies that NED is antecedent of GISE, but not necessarily in a cause and effect relationship.

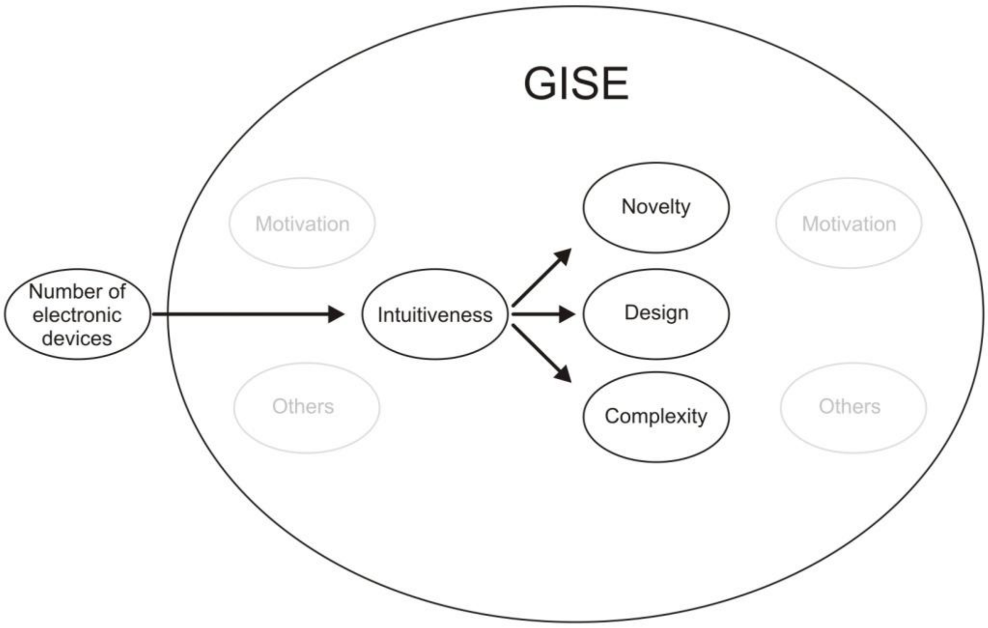


Figure 5-20 Core model

Within GISE, *intuitiveness* is suggested to antecede other latent constructs. Due to theoretical ambiguities, *others* and *motivation* were not positioned within the model at this stage, but it was hypothesized that these may be located either before *intuitiveness* or at the end of the model. Note that the construct *informal help* was named as *others*.

Procedure

The final form of GISE, obtained by the factorial structure revealed after principal component analysis was first trimmed and tested with a first-order path analysis. With these purposes, analyses were conducted on the covariance matrix derived from the final data.

The strategy followed during the procedure was summarized below:

- A covariance matrix consisted of items that are included in the final form of GISE-S, except items that are included in the subscale Formal Help, was derived from the major data;
- The measurement model revealed after principal component analysis was accepted as the first-order model;

The model was trimmed with an aim of having at least 3 indicators that yield high standardized path coefficients for each latent variable, and having acceptable values for the following goodness-of-fit indices[58]:

- Keeping RMSEA and SRMR values below 0.050 for good fit, below 0.080 for reasonable fit (McDonald & Moon-Ho, 2002; Thompson, 2000; also ctd. in Şimşek, 2007; Schumacker & Lomax, 2004; Kline, 2005);
- CFI values above 0.90 (Kline, 2005)

[58] Since there is a lack of consensus in the literature regarding which goodness-of-fit indices should be utilized (e.g. Schumacker, Randall & Lomax, 2004; *Statnotes*, [n.d.]) a relatively large set of indices that are frequently employed where monitored (Schumacker, Randall & Lomax, 2004).

- GFI values above 0.90’s (Raykov & Marcoulides, 2006; Byrne, 1998)
- PGFI values above 0.60 (Stat Notes, [n.d.])
- NFI values above mid 0.90’s (Raykov & Marcoulides, 2006)

Modifications that decrease Chi square / df ratio were looked for (e.g. Statnotes, [n.d.]; Kline, 2005). This ratio was also utilized to compare alternative models in the second stage.

It was ensured that each latent variable was represented by at least three observable variables (e.g. Bollen, 1989; Kline, 2005; Dwivedi et al., 2009).

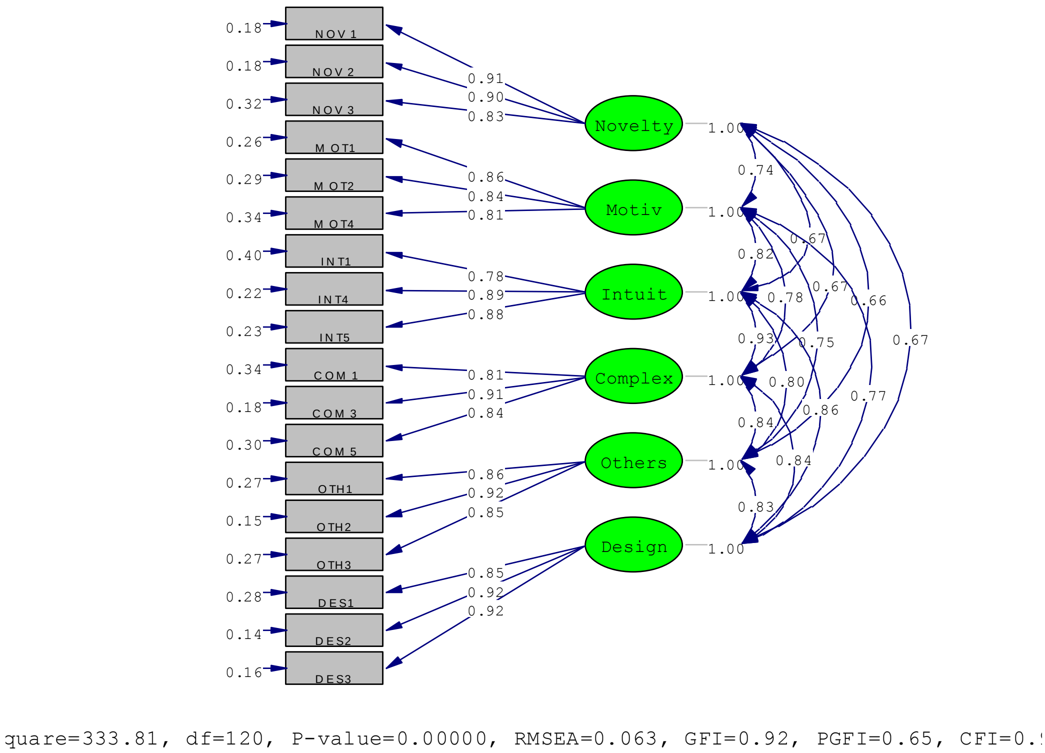


Figure 5-21 Measurement model

A first-order path analysis was conducted for assessing the fit of measurement model using LISREL 8.30 software package. All t values pertaining to paths between latent variables and indicators were significant. After successive reduction of items in order to arrive at a better model, it was possible to retain three indicators for each latent variable and meet goodness-of-fit criteria as well.

All goodness-of-fit indices, except RMSEA were within limits to claim that model fit is good. RMSEA, residing between 0.05 and 0.08 was determined to indicate a ‘reasonable’ fit.

With inclusion of the construct *instruct* (formal help) this trimmed measurement model was suggested as a simplified version (7x3) of GISE-S. This model was briefly presented at the end of this Chapter.

Alternative Models

Before specifying alternative models for exploring and building on the core model presented, variations of the core model were tested.

Structural models for the variations tested are given below:

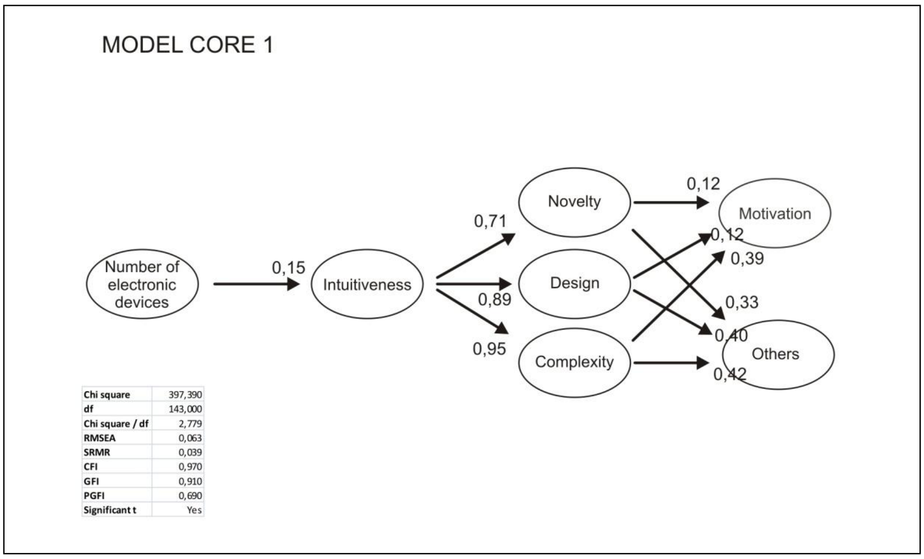


Figure 5-22 Alternative model, core 1

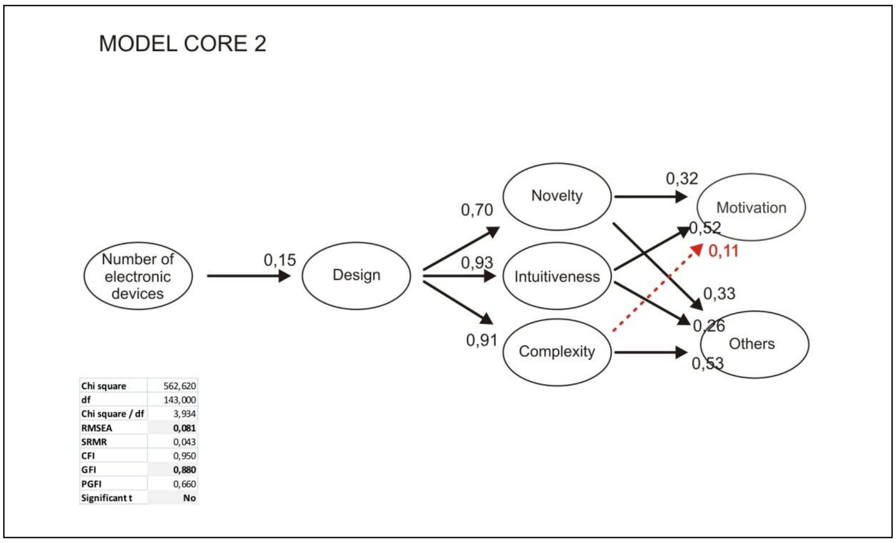
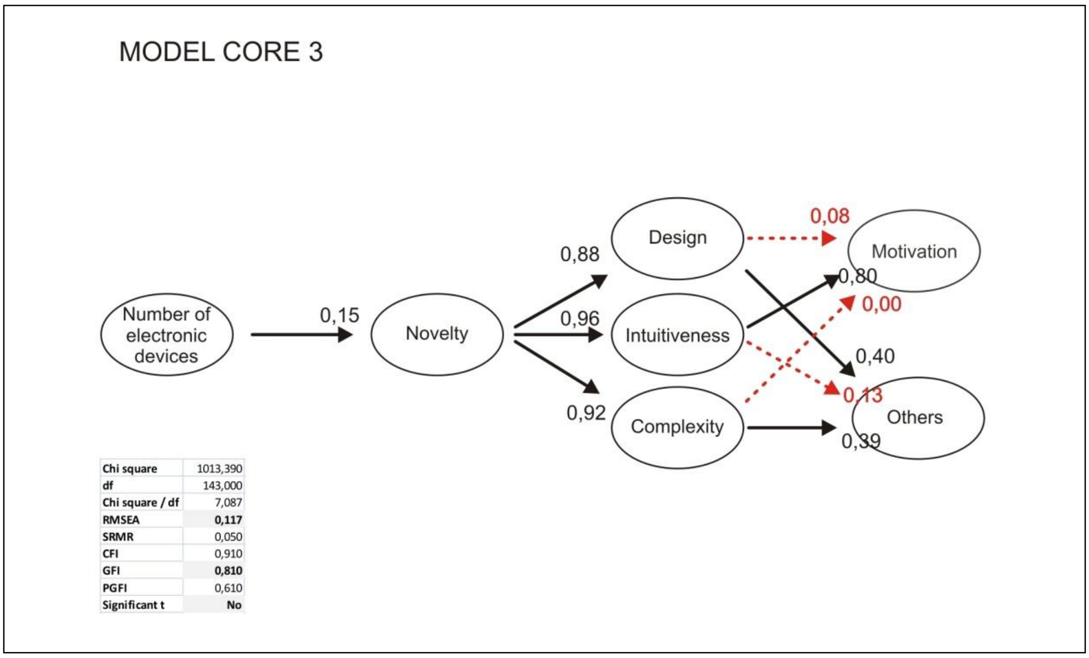


Figure 5-23 Alternative model, core 2



Chi square	1013,390
df	143,000
Chi square / df	7,087
RMSEA	**0,117**
SRMR	0,050
CFI	0,910
GFI	**0,810**
PGFI	0,610
Significant t	**No**

Figure 5-24 Alternative model, core 3

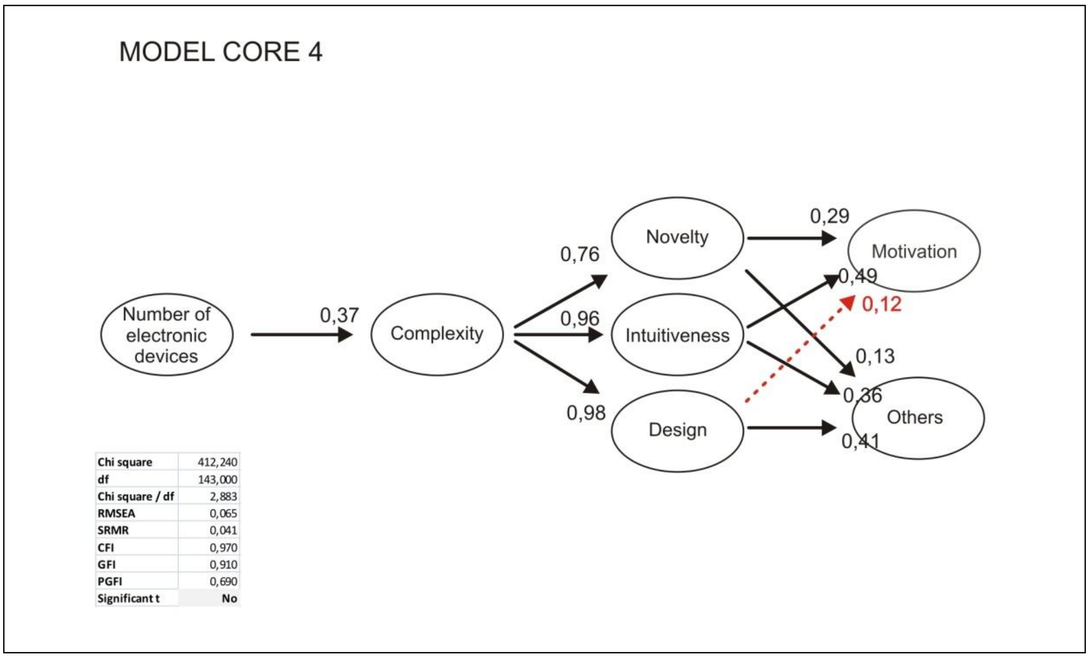


Figure 5-25 Alternative model, core 4

As it can be recalled, Core 1 is the model specified in accordance with the triadic model suggested in this study and the brief theoretical discussion presented before. NED is the only exogenous variable, *intuitiveness, novelty, complexity* and *design* are moderator variables and finally *motivation* and *others* are dependent variables.

Variations were specified in order to find out whether the configuration of the variables *intuitiveness*, *novelty, complexity* and *design* was as it was hypothesized. In other words, the aim was to check whether *intuitiveness* really anteceded other moderator variables or not.

Variations above were constructed in LISREL 8.30 and a path-analysis was done for each. Output files were examined in order to ensure that there were no warning messages. T values for each model were checked to see whether there were insignificant relations. Both standardized estimations and t-values were recorded together with goodness-of-fit indices.

Table 5-24 Goodness-of-fit Indices for alternatives core models

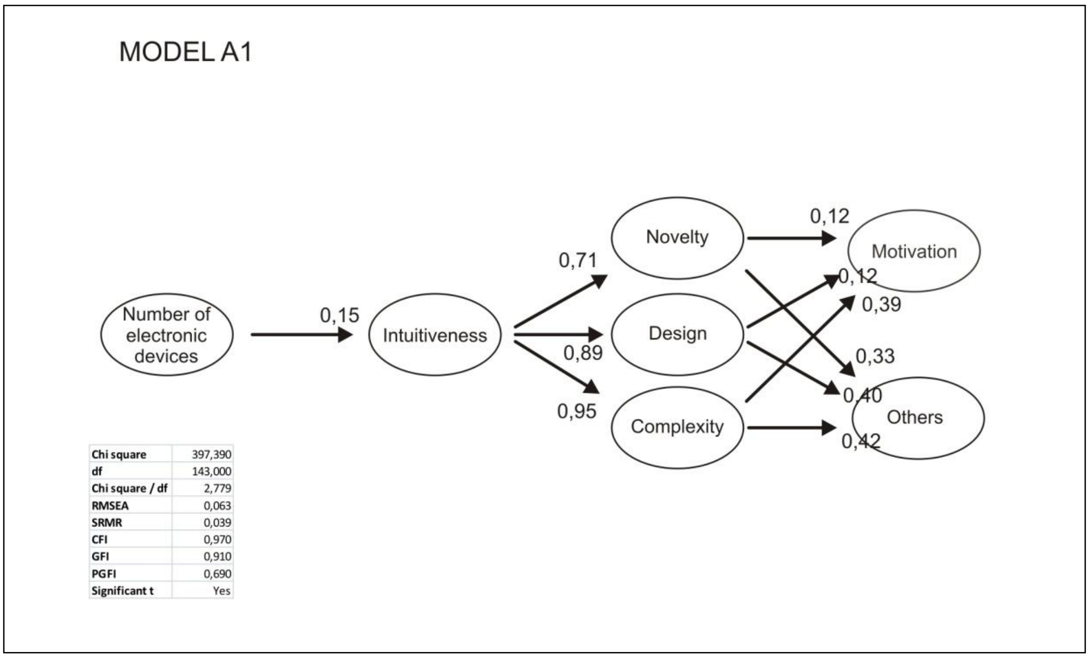
	Core 1 (A1)	Core 2	Core 3	Core 4
Chi square	397,390	562,620	1013,390	412,240
df	143,000	143,000	143,000	143,000
Chi square / df	2,779	3,934	7,087	2,883
RMSEA	0,063	0,081*	0,117*	0,065
SRMR	0,039	0,043	0,050	0,041
CFI	0,970	0,950	0,910	0,970
GFI	0,910	0,880*	0,810*	0,910
PGFI	0,690	0,660	0,610	0,690
Significant t	Yes	No*	No*	No*

* Criterion violated

High RMSEA and low GFI values indicate that models Core 2 and 3 were hard to accept and contain some paths that were not significant (shown with red dashed arrows in Figure 5-23, Figure 5-24) Despite the fact that all t values were not significant, model Core 4 is quite satisfactory as far as goodness-of-fit criteria. It may be speculated that *intuitiveness* and *complexity* are closely related constructs rendering Core 4 a satisfactory model.

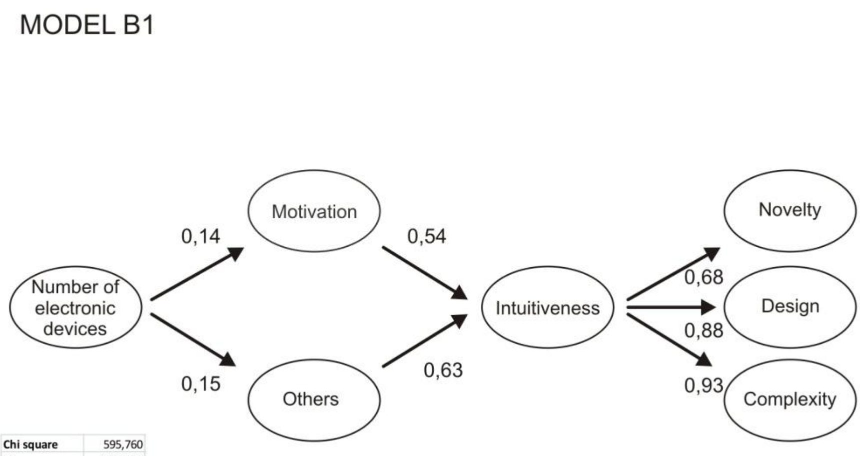
It was concluded that alternative models should be built around Core 1, which yielded best results and is theoretically sound.

Two main alternatives were specified in order to find out whether *motivation* and *others* were dependent variables or not. In this regard model A1 (identical to Core 1) and model B1 were compared.



Chi square	397,390
df	143,000
Chi square / df	2,779
RMSEA	0,063
SRMR	0,039
CFI	0,970
GFI	0,910
PGFI	0,690
Significant t	Yes

Figure 5-26 Alternative model A1



Chi square	595,760
df	146,000
Chi square / df	4,081
RMSEA	**0,083**
SRMR	**0,190**
CFI	0,930
GFI	**0,880**
PGFI	0,670
Significant t	Yes

Figure 5-27 Alternative model B1

Although both models yielded significant t values, goodness-of-fit indices reveal that model A1 fits much better to data. It was observed that model B1 was not able to yield acceptable values for RMSEA, SRMR and GFI. Furthermore, Chi square / df ratio for model B1 almost doubled A1, indicating that model A1 was a superior one.

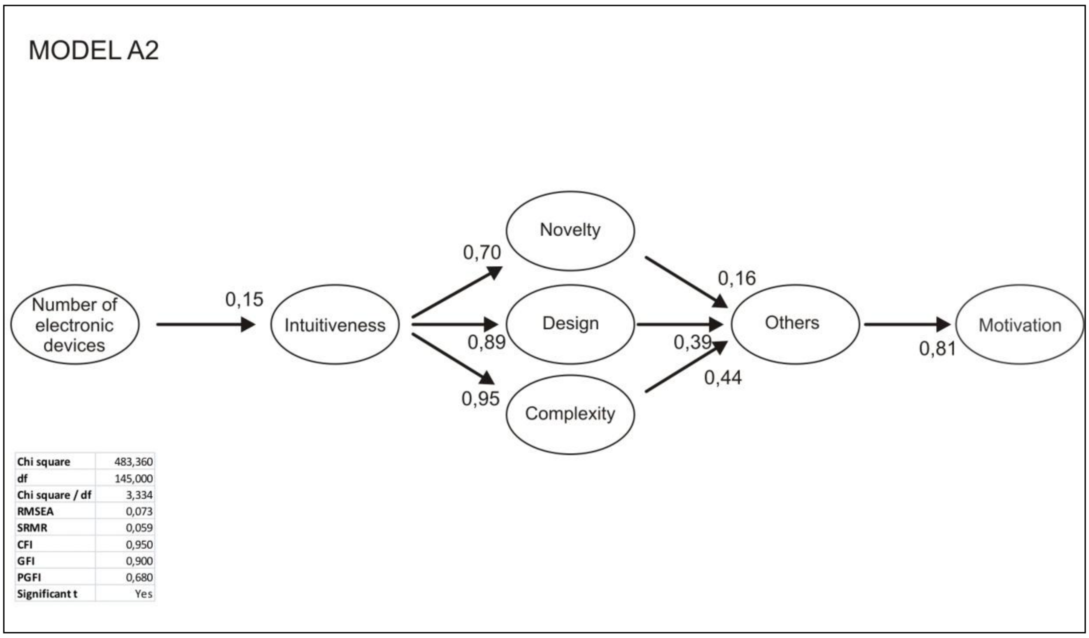
Table 5-25 Goodness-of-fit Indices for models A1 and B1

	A1	B1
Chi square	397,390	595,760
df	143,000	146,000
Chi square / df	2,779	4,081
RMSEA	0,063	0,083*
SRMR	0,039	0,190*
CFI	0,970	0,930
GFI	0,910	0,880*
PGFI	0,690	0,670
Significant t	Yes	Yes

* Criterion violated

Finally, alternative models for model A was specified and tested. Aim of this step was to see whether it was possible to refine model A1 and arrive at a better-fitting structure.

With this aim, 4 alternatives were generated. In A2, *motivation* was specified as the dependent variable, where core constructs were moderated by *others*. A3 was a variation of this where *motivation* and *others* changed places. Finally in A4 and A5 paths between *motivation* and *others* were tested in both directions.



Chi square	483,360
df	145,000
Chi square / df	3,334
RMSEA	0,073
SRMR	0,059
CFI	0,950
GFI	0,900
PGFI	0,680
Significant t	Yes

Figure 5-28 Alternative model A2

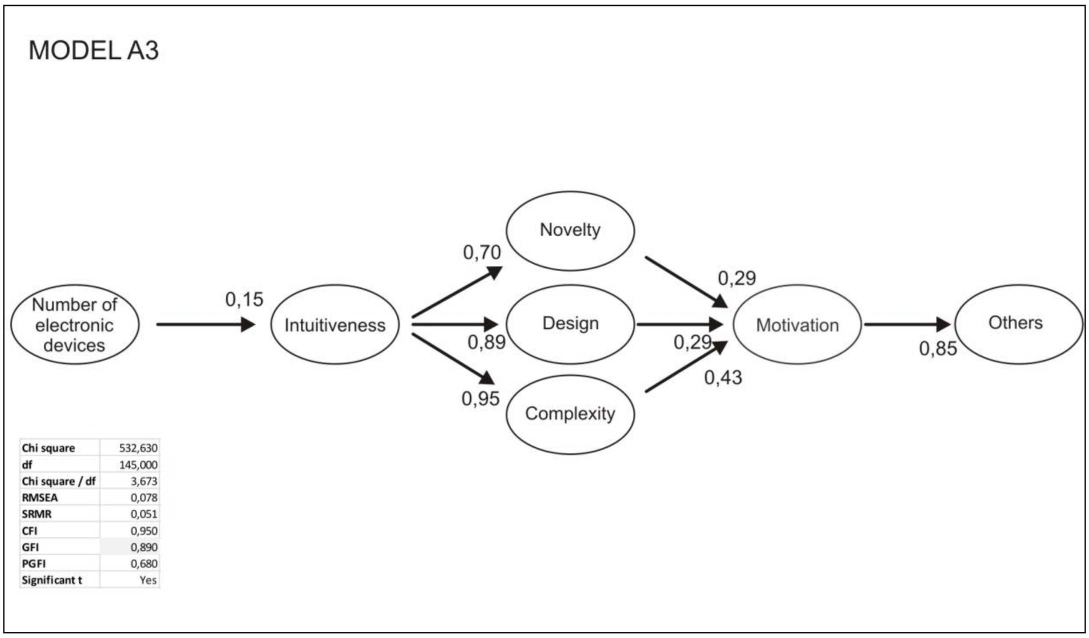


Figure 5-29 Alternative model A3

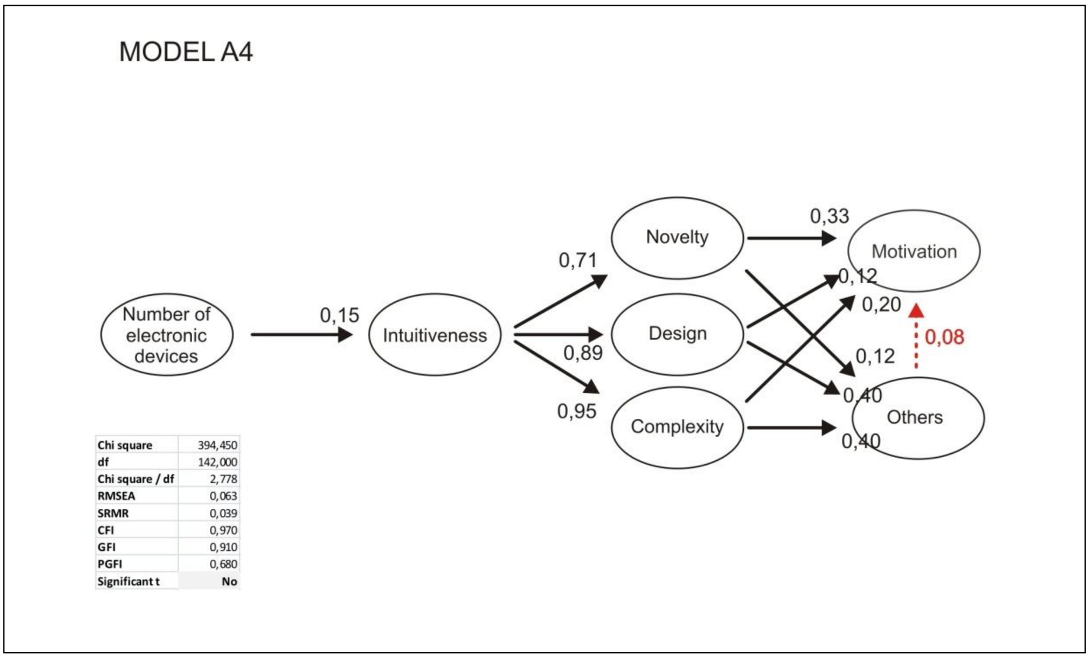


Figure 5-30 Alternative model A4

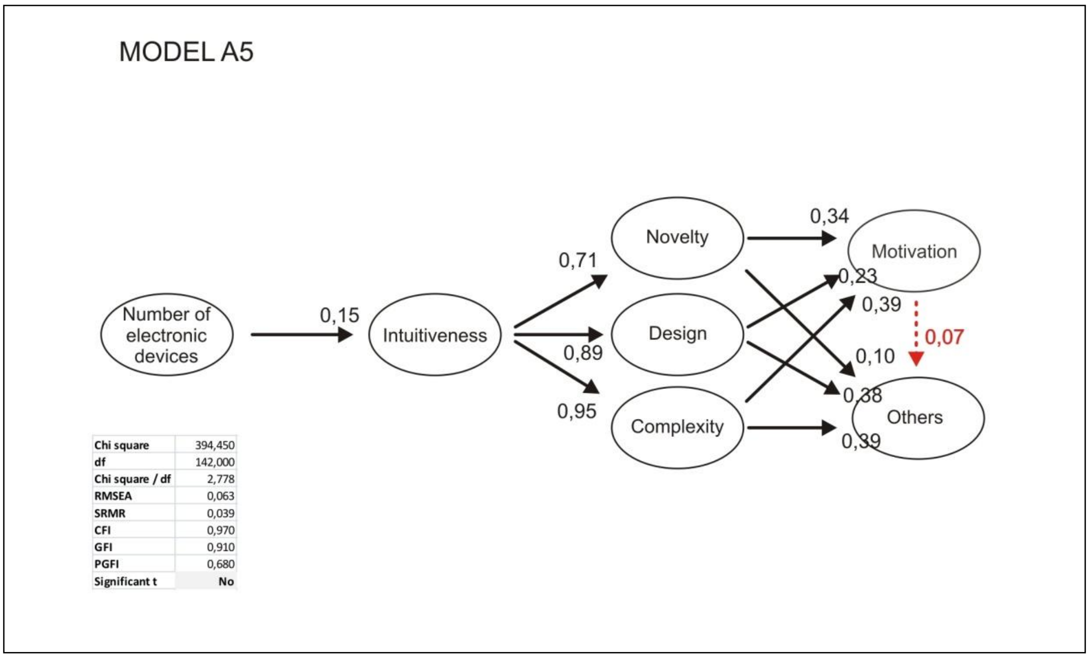


Figure 5-31 Alternative model A5

Analysis of alternatives that explored an additional path between *motivation* and *others*, namely A4 and A5, indicate that paths in neither direction were significant. Results show that models A3 and A4 were acceptable however, they were not better than A1 in fitting the data. Nevertheless, it may be concluded that models that specify *motivation* and *others* as dependent variables, towards the end of the model, fit well.

Table 5-26 Goodness-of-fit Indices for alternatives of model A1

	A1	A2	A3	A4	A5	B1
Chi square	397,390	483,360	532,630	394,450	394,450	595,760
df	143,000	145,000	145,000	142,000	142,000	146,000
Chi square / df	2,779	3,334	3,673	2,778	2,778	4,081
RMSEA	0,063	0,073	0,078	0,063	0,063	0,083*
SRMR	0,039	0,059	0,051	0,039	0,039	0,190*
CFI	0,970	0,950	0,950	0,970	0,970	0,930
GFI	0,910	0,900	0,890*	0,910	0,910	0,880*
PGFI	0,690	0,680	0,680	0,680	0,680	0,670
Significant t	Yes	Yes	Yes	No*	No*	Yes

* Criterion violated

In the light of analyses completed in three steps, model A1 was shown to be acceptable. However, this is not to say there is only one model that is verified by data. It was observed that some of the alternatives of A1, namely A2 and A3, were almost equally acceptable.

It should also be stated that the structural model built in this study neither specifies nor verifies causal relations between the latent variables included.

However, it can be stated that some variables are either directly or indirectly affected by others and some others precede others.

5.10. GISE-S Lite as an outcome of SEM

In the first step of model development process, measurement model was shown to be satisfactory even with a 6 x 3 design. With the inclusion of sub scale *formal help* and eliminating items that yield low path coefficients, a measurement model with a 7 x 3 design was arrived at.

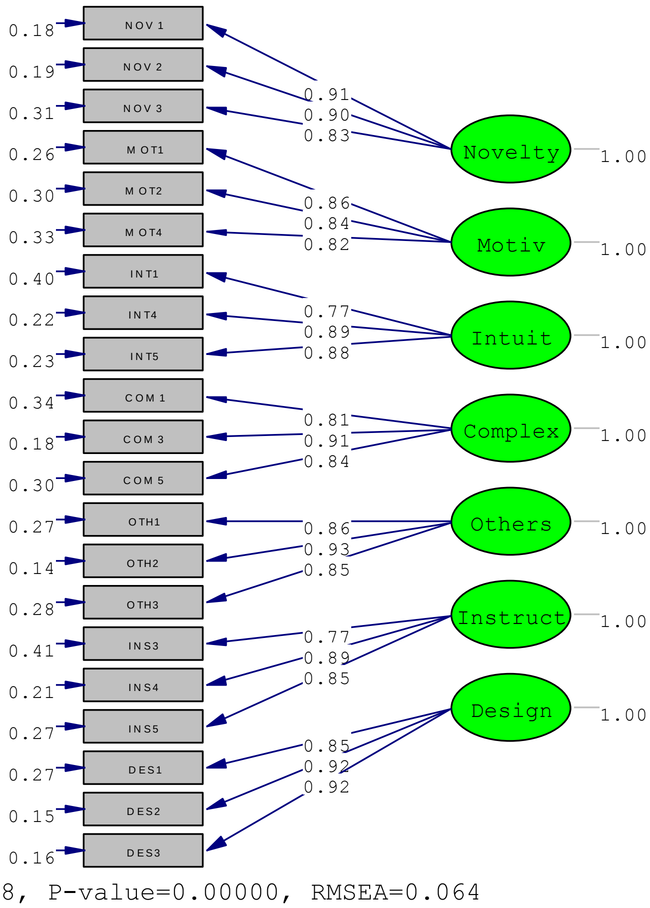


Figure 5-32 Measurement model of GISE-S Lite

GISE-S Lite should be tested in the field, with other samples in order to verify that reliability is not actually sacrificed for the sake of having a more compact design. With further elimination of 14 items, it may be possible to cut down duration of administration by 40%.

CHAPTER 6

6. DISCUSSION: A MULTI-PERSPECTIVE VIEW

As stated in the Introduction, aim of this study was to develop an approach to measure and accommodate individual differences, namely GIE, in usability tests of consumer products. The measurement perspective adopted in this study was to know more about the factors that may obscure the causal link assumed between design and user performance in a test, and to devise cost-effective ways of controlling expertise-related factors quantitatively.

In accordance with this, a nomothetic approach was adopted, that is, rather than trying to explain all that can account for expertise related with the use of digital products in an idiographic fashion, a probabilistic approach was suggested (Babbie, 2001). In accordance with this, prediction with a minimum of predictors rather than a vivid explanation was the ultimate aim. The distinctions between these approaches may best be reflected in the following lines by Babbie (2001):

> The difference between idiographic and nomothetic explanation relates to another distinction [...] [T]he distinction between qualitative and quantitative data. Qualitative data, containing a greater depth of detailed information, lend themselves readily to idiographic explanations. Quantitative data, on the other hand, are more appropriate to nomothetic explanations. Thus, for example, an in-depth interview with one homeless person might yield a full (idiographic) understanding of the reasons for that person's fate, whereas a quantitative

analysis might tell us whether education or gender was a better (nomothetic) predictor of homelessness.

(pp. 74-75)

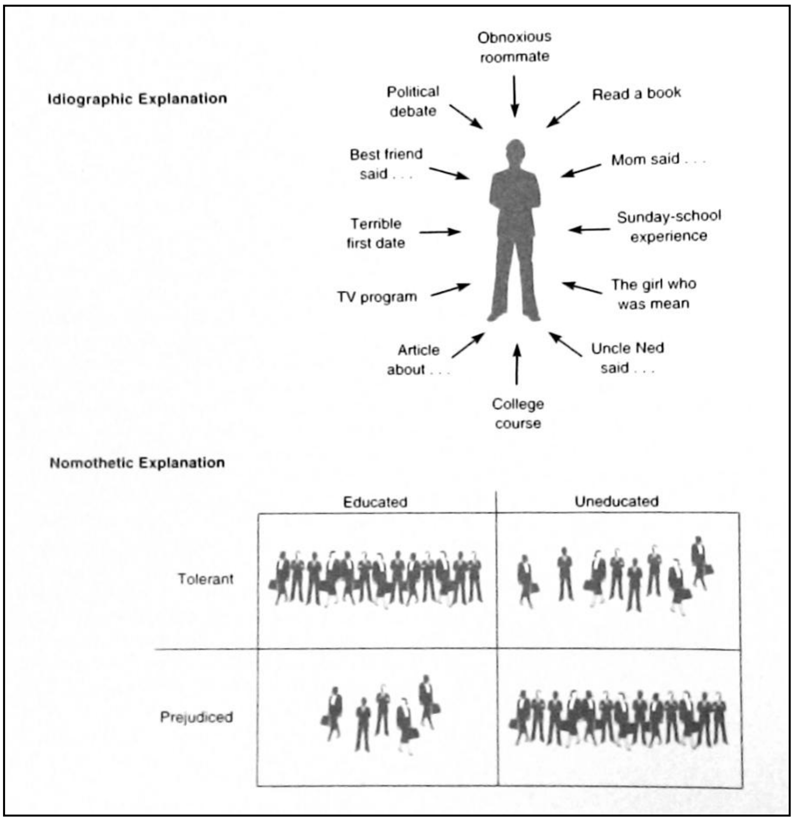


Figure 6-1 Idiographic vs. Nomothetic Explanation [reprinted from E. Babbie, 2001, pp. 74]

Although results and theoretical discussions were treated with a reductionist perspective deliberately, it was evident that a relatively idiographic explanation about phenomena that revolve around GIE and GISE could also be provided. Both perspectives may be regarded as knowing, where measurement may mean 'knowing quantitatively' whereas, qualitative approach may help grasping the plethora of dimensions.

A qualitative approach to the findings may be helpful in non-test situations, where expertise of learning a new device should be studied with qualitative techniques and where it is necessary to gather in-depth knowledge about individuals participated in the study. Especially, in cases where individual accounts of participants should be studied for providing feedback to design decisions and for other generative purposes, outcomes may be utilized as a framework for guiding researchers and designers.

In this Chapter, findings of the study will be discussed encompassing the continuum below.

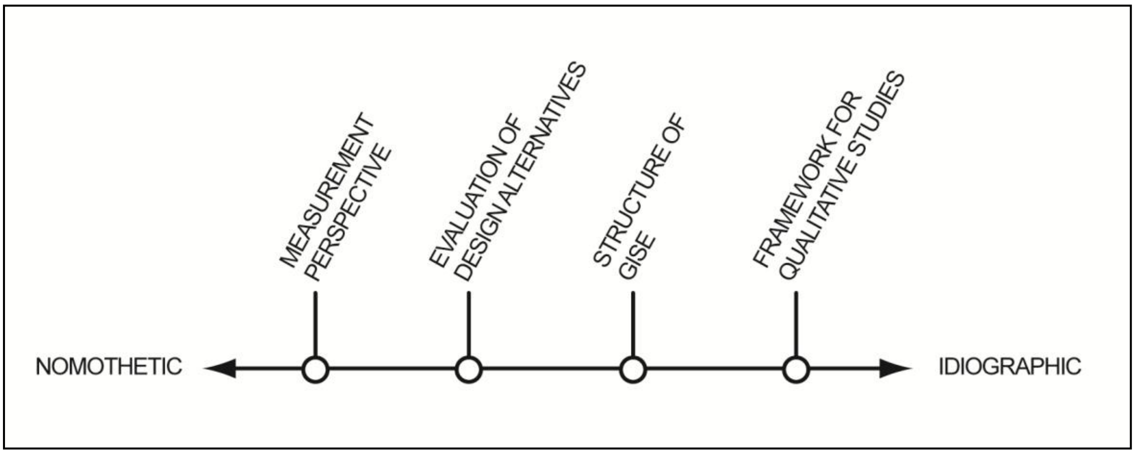


Figure 6-2 Continuum of nomothetic – idiographic approach

In the first part, the results obtained with GIE-T and GISE-S will be discussed; then pros and cons of these two approaches will be compared. In the second part, outcomes of the studies conducted to develop GISE-S will be handled in a different manner and the focus will be on utilization of GISE-S as a means of evaluating design alternatives rather than as a tool for sampling. In the third part, the construct GISE-S will be expanded to reveal its sub constructs and GISE development process will be discussed in the light of SEM results reported in Chapter 5. Finally, the phenomenological model that guided the scale development process will be presented as a framework, and the potentials of this framework as a guide for qualitative studies will be briefly discussed.

6.1. Measurement perspective

In Chapter 4 and 5 the development process, reliability and validity information was provided for both tests. Initial results show that there is prospective evidence indicating that GIE measurement model proposed here may prove to be useful for measurement purposes. In their fully-fledged forms, GIE-T and GISE-S may be valuable tools for sampling or may be administered when any sort of control over experiential factors is necessary.

Depending on the nature of research, tools may be administered in combination or individually, or just in reduced forms. GISE-S, being a paper-based tool, has certain advantages over GIE-T such as cost and ease of administration. However, administration of GIE-T provides the opportunity to observe actual performance of participants. A variety of real-life studies, where tools are administered in parallel to running usability projects are necessary to weigh cost-effectiveness of both tools.

Measurement of GIE may be helpful for:

1) Justification of certain assumptions regarding participant profile;

2) Manipulating GIE as an independent variable;

2) Ascertaining that the effects of GIE on test results were kept to a minimum.

Examples and research scenarios about the potentials of measuring GIE were provided in Chapter 3.

As far as GIE-T is concerned, a further merit of pre-evaluating participants would be to detect the individuals that exhibit intolerable levels of test / performance anxiety before the actual usability test. Furthermore, if normative standards are determined, both tools may also be used to evaluate usability of interfaces in absolute terms. In other words, it would be possible to identify interfaces that require high levels of GIE and those do not.

In the tables below, pros and cons of both tools were listed.

Table 6-1 Pros and Cons of GIE-T and GISE-S

GIE-T
Pros
• Opportunity to observe participant during performance • Face validity is high • Score is available just after test • Since it does not involve attitude measurement, it is not influenced by artifacts such as social desirability or satisficing. • Is a sort of 'standardized' usability test • Shown to have predictive power • Does not seem to cause high 'instrument reactivity'; however, it is a short rehearsal before the actual test—i.e participants may relax after GIE-T and behave naturally

Table 6-1 cont'd
• Behavior during breakdowns and ability to cope with stressful situations are also observed—i.e. Individuals with 'over-sensitivity to being tested' are diagnosed beforehand
Cons
• Time consuming • Tester should be trained • Candidate should be brought to laboratory or to another isolated environment • Requires special software • Some individuals may get exhausted after the test • Content validity is hard to attain • Some participants may feel like a "guinea pig" especially in GIE_PS tasks • Tests should be kept up to date to include state-of-the-art interaction styles
GISE-S
Pros
• Can easily be administered • No need for extra equipment • No need for an isolated environment • Administration in groups is also possible • Easier to integrate to a sampling organization where recruitment agencies

Table 6-1 cont'd
are in charge • Trained testers are not required • Not time consuming, not expensive • Relatively easy to develop – relevant examples and know-how are easily accessed • No need for update, therefore low maintenance costs
Cons
• Needs to be validated and shown that it is reliable • Theoretical basis may be undermined by counter-theories • Inferences may not be straightforward • Intricacies of social sciences should be faced with (especially problems with self assessment) • Can be mistaken for a post-test questionnaire that targets user satisfaction

6.2. Beyond Measurement

6.2.1. Evaluation of Design Alternatives

Up to this point, benefits of measuring GIE were viewed from a measurement perspective. In this section the model will be approached from the other way around and potential uses of the tool as a means for evaluating design alternatives will be discussed. In this regard, findings after the usability tests reported in Chapter 5, for providing evidence for predictive validity will be discussed from another perspective. As it may be recalled, in both tests it was shown that GISE-S values were highly correlated with usability test results, but there was a 0.34 point difference between the correlation coefficients.

If the definition of GISE is revisited one may generate ideas in order to explain the 0.34 point difference between the studies. In Chapter 2 GIE was defined as follows.

> **General Interaction Expertise (GIE)** is acquired by experiencing several interfaces and helps users to cope with novel interaction situations.

Commencing with this definition, GISE was defined as follows:

> General Interaction Self-Efficacy (GISE) is a judgment of capability to establish interaction with a new device and to adapt to novel interaction situations…

As it can be seen, GISE was defined as a construct to denote the changes in individual's attitudes towards her or himself, induced by several positive or negative cases of interaction. In this sense both GIE and GISE may be briefly defined as adaptations in order to cope with novel and unfavorable situations. It is evident that users exhibit individual differences with regards to 'ability'[59] to cope

[59] The term 'ability' is not used to denote a basic cognitive ability.

with unfavorable conditions, and in turn some of them perform well, while others experience problems. Although this argument holds true in many cases, one of the essential factors may be missing in some circumstances rendering this correlation useless.

6.2.2. Design characteristics: Link between GIE and Usability Performance

While relating GIE with usability performance, there is a crucial moderator which makes this link possible that is design. From design perspective, ideally an interface should make it possible for everyone to have a problem-free experience. In ideal conditions, there should be no correlation between GIE and usability performance. However, it should be noted that there may be no correlation between GIE and usability performance when the interface is almost impossible to use for even the most experienced users. In other words, in cases where design is so successful that everybody may sustain a problem-free interaction GIE should play no role. This observation will also be valid for cases where design is so poor that nobody is able to use the product.

Within this perspective, measurement of GIE, either with GIE-T or GISE-S may enable designers and researchers to compare two interfaces and determine the one that requires less GIE, or that is more intuitive.

In Study 2 and 3 presented in Chapter 5, two products were tested and GISE-S was administered to participants. Since no actions were taken against, mean and dispersion of GISE-S scores were not the same for two studies and participant profile exhibited variation with regards to GISE-S. If descriptive statistics calculated with data gathered in major data collection phase are assumed as normative, mean GISE-S z-scores in Study 2 and 3 would be +0.45 and +0.85 respectively. In other words both samples were positively biased with regards to GISE, where individuals participated to Study 2 were almost one standard deviation above the population mean[60], whereas participants of Study 3 were 0.5 standard deviation above the population mean.

As far as usability performances are concerned, participants in Study 3 were more successful (0.56) than the ones in Study 2 (0.50).

If GISE-S is accepted as a reliable and valid scale then it may argued that product tested in Study 3 (an IPTV) had a better interface design regarding usability than the cellular phone tested in Study 2. This result is also in line with the fact that although a very high correlation was observed between GISE-S scores and usability performance for the cellular phone (r=0.95), this was not the case for the IPTV (r=0.61).

It should be noted that usability performance—i.e. effectiveness scores, is not only determined by design characteristics, but also by other factors that delineate what is experienced by participants. Tasks selected, the way test was conducted, timeout thresholds and some others affect what is experienced by the participants.

In order to put the phenomenon technically more accurate, terminology should be clarified and the relations should be simply defined.

GIE level: General Interaction Expertise of participants

[60] Actually, the sample size in major data collection phase is far from representing the population. Here this data was utilized for comparing samples in Study 2 and 3.

Experience Difficulty: Test difficulty that is determined by design characteristics, complexity of scenarios, whether time limits are set for scenarios, assistance provided during tests, and all the other factors that may alter effectiveness scores

Usability Performance: Aggregate effectiveness scores for each participant across all scenarios included in the test.

It may be assumed that if Pearson's r between GIE and usability performance is low but usability performance is high (see quadrant III in Figure 6-3) the experience difficulty is extremely low. If r is low but usability performance is also low (see quadrant IV in Figure 6-3) then it may be concluded that Experience Difficulty is extremely high.

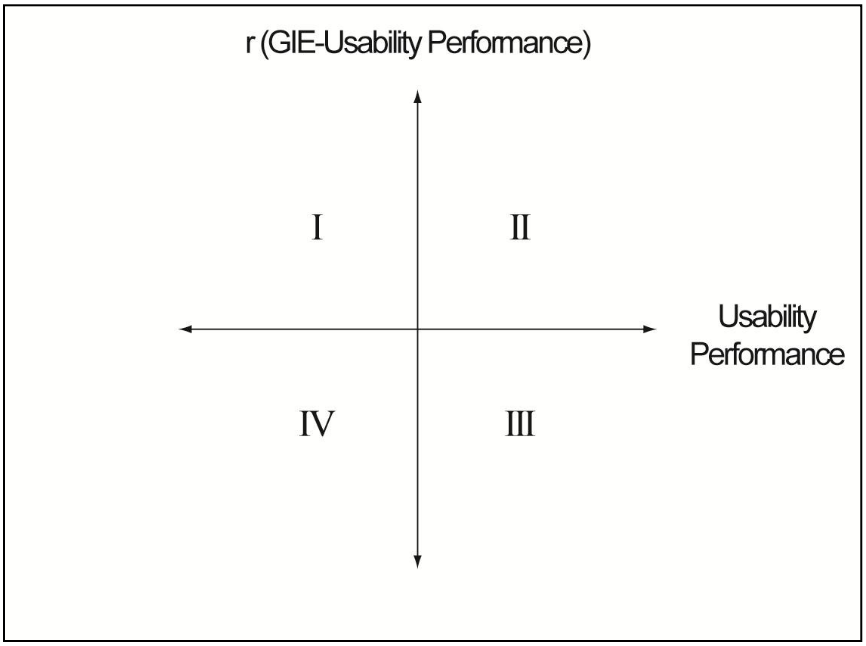


Figure 6-3 Relationship between r (GIE-Usability performance) and usability performance

It should be noted that these interpretations may only be valid if average GIE levels of participants reside around the population mean. If GIE levels are extremely low or high, or variance is too low (for example if GISE-S scores are in the range of 100 ± 5) these relations may no longer be valid. Moreover, factors other than design characteristics should be isolated to augment the effect of design on the results, so that alternative designs may safely be compared.

Going one step further, it may be argued that the correlation of subscale scores with Usability Performance may also be interpreted in certain ways. If the correlation between individual subscale scores and usability performance scores were compared, it can be seen that all the subscales yield high and significant correlation coefficients in Study 2 (see 5.8.1). However, in Study 3 (see 5.8.2) *formal help*, *specific design characteristics* (design), *motivation* and *informal help* (others) scores correlated significantly with Usability Performance. Although, it is interesting to see that some of the subscales correlated well while other did not, interpretation of this finding at this stage is not an easy task.

With additional studies that are experimental in nature, how certain interfaces "tap" certain sub constructs should be explored in order to look for patterns that may give valuable information for designing easy-to-use interfaces or generating user profiles like personas (Cooper & Reimann, 2003).

In such studies, certain patterns or 'personalities' may be associated with certain behavior or preferences. For example, users that rely on others to learn and have low self-efficacy regarding learning novel interfaces may be explored compared with self-learners who enjoy experiencing novel interfaces regarding expectations from a new interface.

Findings up to this point indicate that measuring GIE is not only useful for controlling individual differences in usability tests, but also for exploring to what extent certain interfaces or parts of interfaces tap GIE.

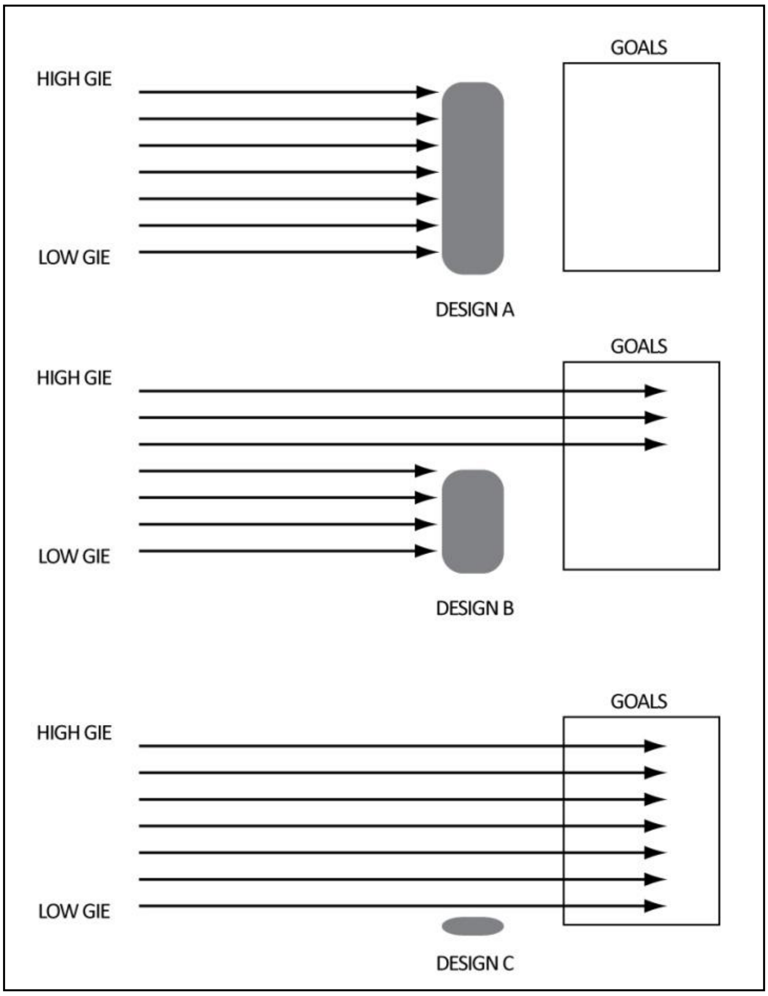


Figure 6-4 Relationship between GIE, design characteristics and accomplishing goals.

Within this approach both GIE-T and GISE-S may be employed to compare design alternatives, different modes of interaction or individual features and scenarios of a particular product.

Furthermore, GIE-T or GISE-S may be partially administered in order to see how certain behaviors (in the case of GIE-T) or sub constructs (in the case of GISE-S) interact with certain design alternatives or features.

In addition to this, individual sub scale scores may be utilized as a means of user profiling, where GISE-S is administered to a large sample, and handled with a multi dimensional approach.

6.2.3. Structure of GISE

As a second outcome of the validity studies conducted in this project, structural relations within GISE was specified with a model built with SEM technique.

In this section, the construct of GISE will be expanded first for discussing the structural model built in Chapter 4. In this discussion GISE will be handled in a different way to bridge the gap between nomothetic and idiographic approaches briefly presented in this chapter.

As users experience digital[61] products they have both positive and negative experiences about them. Before acquiring a certain amount of GIE, users prefer and use products with intuitive interfaces. This behavior may be exemplified by users looking for simple interfaces and even sacrificing functionality. Avoiding complex functions of a product and using only some basic features may also be associated with behavior that users with low GISE would exhibit. Such individuals may get frustrated in situations when they had to learn new products. Such circumstances may be irresistible when user had to replace a product which is

[61] Note that the term “electronic device” in NED was suggested for the sake of clarity while administering LEDQ.

indispensible for them (e.g. a cellular phone) or others decided to renew a product that was in joint use (e.g. a television set, or a new alarm system). Motivation by necessity (i.e. usefulness) and lack of negative feelings may be crucial for them, together with help from others to support them while they learn the new product (see 1 in Figure 6-5).

As users gain a certain amount of GIE and further build GISE beliefs, they may try mastering non-intuitive interfaces and attempt to manage complex, novel products that do not comply with good interface principles (see 2 in Figure 6-5). Users may be more willing to attempt to learn a new product at this stage even if they are not necessary to do since the cost of learning is not so high for them. With new experiences they would either strengthen their GISE or lose confidence.

At this level, good performers would rely less on others' help and non-intuitive products would no more pose a problem for them. Ultimately, as their GISE beliefs get stronger they would be confident in learning new and complex devices on their own and even start to help others. Eventually, they would start to enjoy learning process. This would help them build an even stronger GISE, and together with the help of other transformations they will believe that they can easily learn a new product even if they are not motivated by usefulness or affection (see 3 in Figure 6-5).

Soon, they would start to get involved into more learning situations in their jobs and family life owing to their strong GISE (see 4 in Figure 6-5) and their expertise will turn into a social role. It is even claimed that such individuals are known to choose, configure or customize digital products so that perceived complexity is increased to underscore their expertise even stronger (Kiesler et al, 2000).

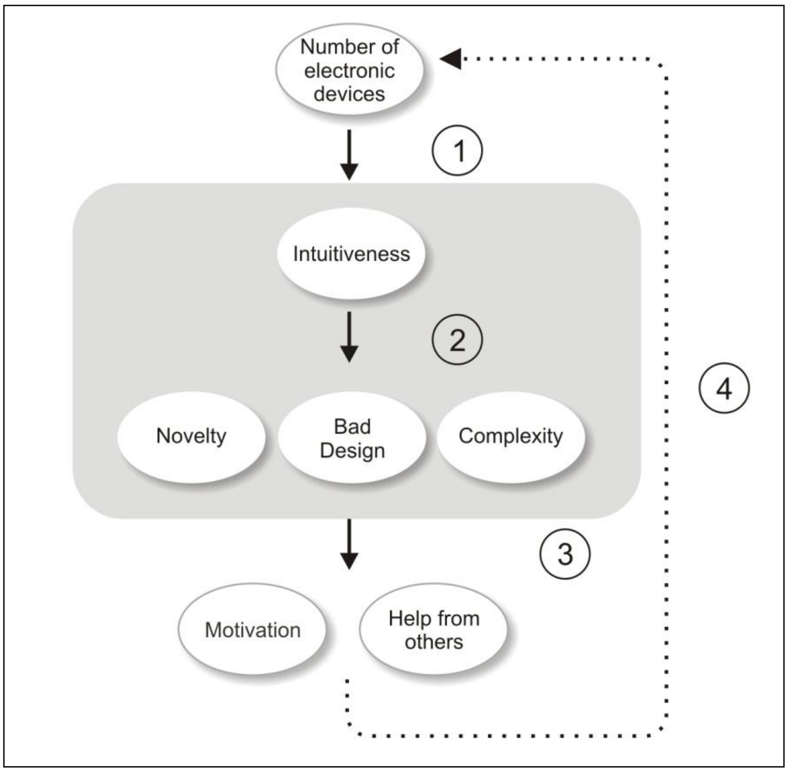


Figure 6-5 Structure of GISE

In that sense intuitiveness is not a requirement for them. It may even be argued that such users may start to look for highly complex systems where ease-of-use is not a concern, or sacrificed for reducing costs or for more functionality. This may be exemplified by a computer enthusiast who rejects using systems with a graphical user interface and insists on programs that utilize command based interfaces.

6.2.4. A framework for Qualitative Studies

As mentioned in Chapter 5, the primary source for item pool was 550 negative and positive expressions that respondents subjectively gauge their self-efficacy beliefs. The vividness of the original phenomenological model was partially reflected in the final form of GISE-S and the structural model.

The opportunities of using the phenomenological model developed with the results of LEDQ as a framework was not discussed in a detailed fashion. This phenomenological model, together with the structural model discussed here may be utilized for studying individuals' personal histories or styles of developing GISE during the acquisition of GIE. Furthermore, framework may prove to be useful if employed in order to study what individuals experience during learning a new digital product (i.e. while acquiring SS; see Chapter 3) or a new family of products (i.e. while acquiring a specific AS).

In qualitative research, even when data is collected with unstructured interviews, it is devised that a framework called 'aide-mémoire' is established in order to guide the process (e.g. Briggs, 2000; Zhang, 2006). These agenda serve as guides so that every aspect of the phenomenon are discussed and individual interviews are kept in a definite scope, rather than a specific list of questions to be asked (Zhang, 2006). The phenomenological model presented in this study (see Figure 5-9) can be utilized as a general *aide-mémoire* to explore several aspects of GIE – GISE related constructs. Furthermore, the model may be utilized as a template for affinity diagrams or visual databases where data is sorted or to track data collection process so that researchers might decide whether saturation occurred and study should be terminated or not.

With the speculative scenarios below, how this model may operationally be used in several settings was tried to be illustrated.

It was left to researchers to translate LEDQ expressions that form the atomic elements of the phenomenological model into mini tour questions and categorize them to obtain grand tour questions (Spradley, 1979).

Research scenario I

In a field study, a prototype trial is going to be carried out in order to explore the reactions of a diversity of participants. Researchers decide to see how different individuals succeed or fail to build self-efficacy with regards to a novel product. In this case the model may be used as an aide-mémoire to capture the experiences of individuals during successive home visits.

Research scenario II

In a participative design study of a new product, in order to include extremes into the study, individuals are interviewed to learn about their personal histories and styles of learning to use a specific family of digital products. Individuals are grouped into a set of classes reflecting their styles, instead of their expertise levels, and feedbacks they provided are interpreted in accordance with their styles and choices.

Research Scenario III

In a comparative study, participants are given enough time to experience and learn to use two alternative prototypes. User experiences in the process of learning of both prototypes are compared by a post-study interview, based on grand and mini tour questions derived from the model provided.

Research scenario IV

In a prototype trial, a new product is given out and the learning process is monitored with a longitudinal study. In certain periods, home visits are carried out and problems witnessed are organized with the model provided in the form of a conceptual map.

CHAPTER 7

7. CONCLUSION

In this chapter, first a brief review of answers acquired during research, based on literature review and empirical studies will be presented.

In the second part, an integrated model will be presented that schematizes all the constructs studied and combines partial models utilized throughout the study into a single conceptual model. A concise meta-discussion of the work done in this study will be done with reference to this model.

In the third part, limitations of the study will be discussed. Finally, further studies that are required to complement the progress made will be suggested.

7.1. Answers acquired

As the reader may recall, research questions were addressed in the Introduction, with an aim of first defining the problem, and then devising ways for studying the problem. The primary aim of the study was stated as follows:

> “...to develop a framework to accommodate experiential factors in usability tests and other user-centered design techniques in the case of consumer products, so that results are not affected by individual differences.”

In order to attain this aim, the following questions were tried to be answered during research.

7.1.1. What is mainstream approach to sampling in usability studies?

Before defining the problem, it was stated that problem with testing of consumer products was the application of conventions valid for the domain of HCI to the domain of consumer products in a verbatim fashion. In accordance with this, it was suggested that homogeneity assumptions valid for professional products may not be valid in case of consumer products. Then literature was revisited to see whether mainstream approach in sampling was suitable for testing consumer products. Through the literature review, it was observed that current approach to sampling was rather problematic in the way that experiential factors are treated. The common practice was determined as utilization of readily observable variables to represent experience.

7.1.2. What are the individual differences that may affect usability test results? Do experiential factors play a significant role?

Several types of individual differences that may affect usability test results were enumerated in Chapter 2. Literature findings emphasized the significance of experiential factors, which was actually rationale behind the study. It was found that experiential factors were listed among the most important factors to be considered during sampling by many authors. However, a proper way of handling these factors was not recommended.

7.1.3. How should experiential factors be approached so that they no more obscure the link between design characteristics and usability performance?

It was concluded that it is not plausible to reduce experiential factors to what was experienced by the individual. Although experiential factors are influenced by what was experienced, it was argued that the changes induced should be focused on. Therefore, an approach based on “expertise” was adopted. With such a perspective, expertise was defined as an attribute that influences performance directly. However, reservation was left for other variables such as gender, age, education level and others. After empirical studies, it was shown that those readily observable variables may correlate with experiential factors. Neverhtheless, this relation is most probably indirect—i.e. moderated by the quality and quantity of experience with digital products.

In the rest of the study, the main effort was to measure "expertise" in different ways so that a triangulation was possible, as well as alternative tools to be employed under a diversity of circumstances.

It may be concluded that in order to maintain that the link between design characteristics and usability performance is visible, controlling experiential factors are necessary. The nature of control may vary depending on the research design. For example, experiential factors may be measured for screening purposes and ensuring that several samples are comparable with regards to expertise. In another research setting, measurement may be utilized for handling level of expertise as a treatment group. Regardless of the way it is employed, measurement should be done for transforming experiential factors to a variable that enhances research designs rather than inducing systematic error.

7.1.4. How can experiential factors be approached within a measurement perspective?

Within a measurement perspective, first a construct definition (GIE) was developed to guide the whole process. Then, concrete manifestations of this construct were looked for. With this aim, based on Bandura's Social Learning Theory (see Chapter 3), a triadic model was proposed to specify how people acquire GIE and the transformations took place during this process. This main model was augmented with additional models, and then, with empirical findings (see Chapter 4 and 5).

It was argued that, GIE was a latent construct by definition, and could only be 'observed' indirectly through its reflection in certain mechanisms. Based on the triadic model, a two-fold measurement scheme was proposed that target both actual performance (GIE-T) and attitudes (GISE-S).

Measurement of actual performance was formulated as a straightforward tool, where automatic and controlled processes were targeted by individual apparati (GIE_XEC and GIE_PS). In order to grasp attitudes that reflect and moderate performance, a construct called General Interaction Self Efficacy was defined. A scale to measure this construct was developed. Reliability and validity evidence was provided for each tool. However, additional studies are necessary.

7.1.5. How can this framework be utilized for evaluating design alternatives?

Although tools that target GIE may be regarded as valuable additions to researcher's and designer's toolbox, a further means of utilizing this was suggested. It was stated that ideally a design should be easily used by everyone, and expertise should not play a role in enhancing one's performance. Stemming from this assumption, measurement of GIE may be suggested as a benchmark against which design alternatives may be compared (see Chapter 6).

7.1.6. How can this framework be utilized in qualitative research?

In this study a research strategy based on convergence was employed. Although primary aim was to handle phenomenon in a minimal fashion so that measurement was possible, at early stages phenomena targeted were broadly defined and their plethora was tried to be grasped. At later stages this richness was sacrificed for the sake of parsimony through controlled processes of reduction. While this reduction process enabled to establish a measurement framework, it was thought that initial findings could serve as a road map whenever plethora of dimensions should be studied.

The phenomenological model derived from respondents' ideas about favorable and unfavorable conditions when learning a new electronic device may be defined as a plethora of dimensions of this sort. This model, together with the structural model built with SEM technique, may serve as an *aide-mémoire* while conducting qualitative studies. Furthermore, the phenomenological model may be developed to aggrandize differences and define axes on to which users may be mapped to define patterns, as in the case of developing personas.

7.2. Integrated model

The model that integrates all the partial models suggested in this study is presented in Figure 7.1. As it can be seen, the main relation explored in this study was the one between experience and usability performance.

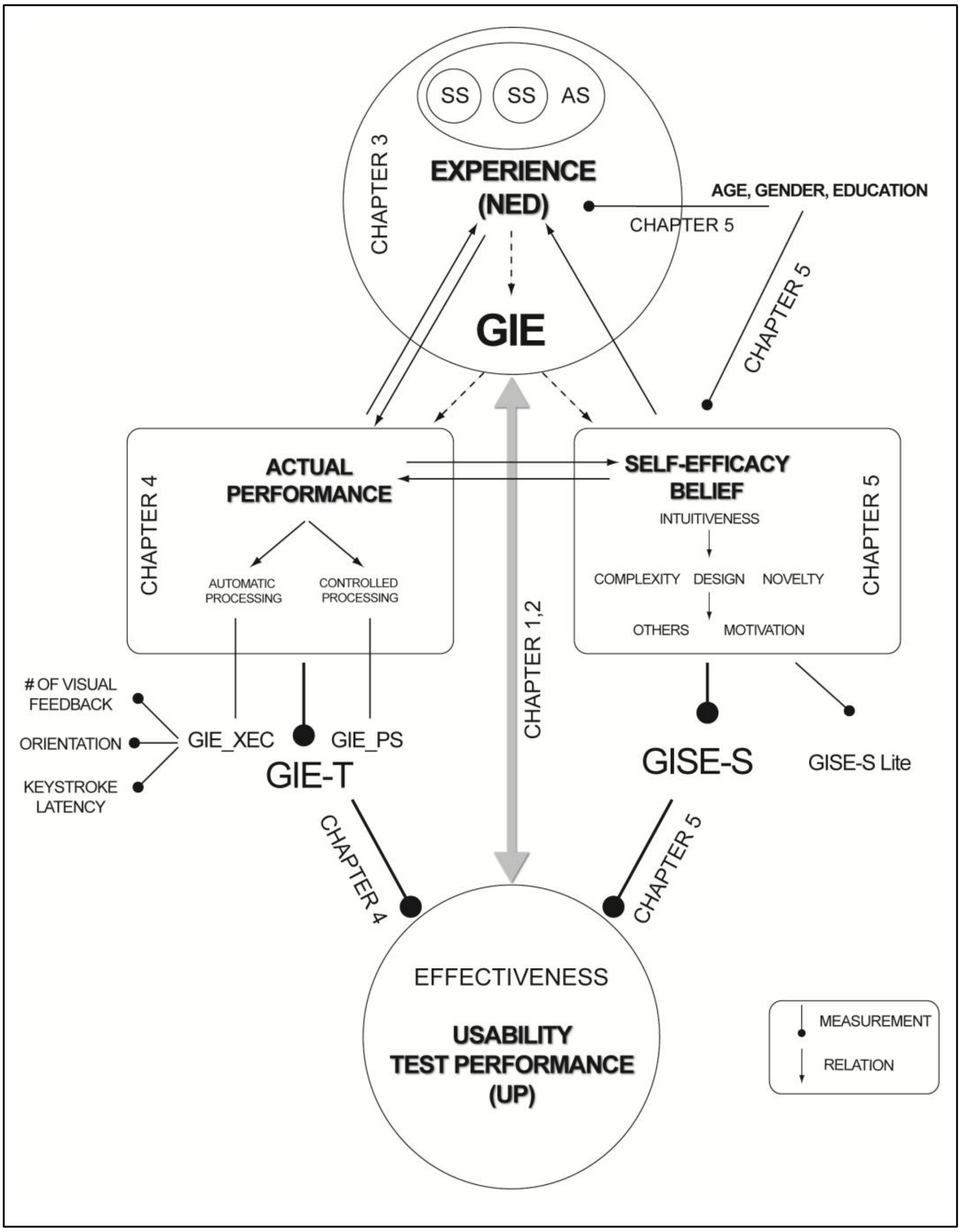


Figure 7-1 Models Integrated

As it was put forward in the theoretical discussions throughout the study, since GIE is a latent construct, this relation was assumed to be moderated by actual performance and attitudes. These were depicted as two main paths that link experience and usability performance.

The integrated model consists of the experience model presented in Chapter 4 (see Figure 3-3), the triadic model (see Figure 3-1) and finally the structural model developed with SEM (see Figure 5-26).

In addition to these, some auxiliary findings were tried to be explicitly put in this model. For example, an alternative to GIE_XEC score was found out to be # of visual feedbacks, orientation or various types of keystroke latencies. These measures may be worked on as to devise an easier and cheaper way of observing actual performance.

Similar to that, the effect of gender, age and education, which were discussed in Chapter 5, were included to form another triadic relationship between NED and GISE.

As it can be seen in the integrated model, the link that was not studied in any means was between experience and actual performance, and the work was concentrated mostly on the GISE path. This was mainly because the fact that working on GIE-T was more time consuming and it was only possible to develop GIE-T as a 'proof of concept'. GISE-S, on the other hand, was almost fully developed, together with a 'lite' form to further reduce administration costs. Nevertheless, theoretical framework for GIE-T that is based on the dichotomy of controlled vs. automatic processing can be defined as a parsimonious and firm framework, which is in line with main learning or skill acquisition theories that pertain to schools of information processing and activity theory.

7.3. Limitations of the study

Although almost all research questions were answered, there were certain limitations of the study.

As it was previously mentioned, due to its costly nature, it was not possible to develop GIE-T into a fully-fledged tool. In this regard, GIE-T may be regarded as a prototypic tool, or a proof of concept. Especially, in the case of GISE_PS, it was only possible to show that such apparatus tests would be valuable in targeting controlled processes.

Second, it was not possible to administer both tools in real-life settings to see how they interact and how they correlate. Validity studies were conducted separately and there were no opportunities to observe whether it is possible to augment the predictive power when tools are administered in combination.

Another limitation was the fact that reliability and factor structure was not tested with a new sample, although scale was administered to small sets of participants.

7.4. Further studies

Further studies are necessary in order to obtain a full proven measurement framework and fully-fledged tools.

GISE-S should be translated to English using specific techniques to guarantee accuracy. Having an English version of GISE-S is necessary for dissemination of knowledge and for exploring intercultural aspects with regards to GIE. For these purposes, GISE should be administered to a sample in English and results should be compared.

Data should be collected with GISE-S or GISE-S Lite in order to provide further information on reliability and validity of the scale. In this regard, known groups comparison and questionnaires that may open up opportunities to situate GISE on a nomological network may be worked for.

New items and parallel forms should be developed and prototyped especially for GIE_PS, in order to have a tool that can be administered in real-life situations.

The phenomenological model specified after LEDQ and the structural model built with SEM technique should be explored qualitative through interviews and field studies in order to gain more insight so that social and cultural aspects are studied as well.

Furthermore, experimental research is necessary for studying how this measurement framework may be utilized for comparing design alternatives and understanding constructs defined here.

REFERENCES

Ackerman, P. L. (1987). Individual differences in skill learning: An integrating of psychometric and information processing perspectives. *Psychological Bulletin , 102* (1), 3-27.

Ackerman, P. L., & Humphreys, L. G. (1990). Individual differences theory in Industrial and Organizational Psychology. In M. D. Dunette, & L. M. Hough, *Handbook of Industrial and Organizational Psychology* (2nd edition ed., pp. 223-283). California: Consulting Psychologists Press.

Adler, P., & Winograd, T. (1992). *Usability: Turning technologies into tools.* New York: Oxford University Press.

Aiken, L. (2000). *Psychological testing and assessment.* Boston: Allyn and Bacon.

Anastasi, A., & Urbina, S. (1997). *Psychological Testing.* New Jersey: Prentice Hall.

Babbie, E. (2001). *The practice of social research.* Belmont, CA: Wadsworth/Thomson.

Bandura, A. (1986). *Social foundations of thought and action.* London: Prentice .

Barbeite, F. G., & Weiss, E. M. (2004). Computer self-efficacy and anxiety scales for an internet sample: testing measurement equivalence of existing measures and development of new scales. *Computers in human behavior , 20*, 1-15.

Benbasat, J., Dexter, A., & Masulis, P. (1981). An experimental study of the human / computer interface. *Communications of the ACM* , 752-762.

Berkman, A. E., & Erbuğ, Ç. (2005). Accommodating individual differences in usability studies on consumer products. *11th conference on human computer interaction, 3.*

Bodker, S. (1991). *Through the interface. .* Lawrence Erlbaum: Hillsdale.

Bollen, K. (1989). *Structural equations with latent variables.* New York: John Wiley.

Bong, M. (2006). Asking the right question: how confident are you that you could successfully perform these tasks? In F. Pajares, & T. Urdan, *Self-efficacy beliefs of Adolescent* (pp. 287-307). Connecticut: Information age.

Briggs, C. (2000). Interview. *Journal of Linguistic Anthropology* , 137-140.

Bunz, U. (2004). The computer-email-web (CEW) fluency scale—development and validation. *International Journal of Human-Computer Interaction , 17* (4), 479-506.

Bunz, U., Curry, C., & Voon, W. (2007). Perceived versus actual computer-email-web fluency. *Computers in Human Behavior , 23*, 2321-2344.

Byrne, M. (1998). *Structural Equation Modeling With LISREL, PRELIS, and SIMPLIS.* New Jersey: Lawrence Erlbaum.

Card, S., Moran, T., & Newell, A. (1980). The keystroke-level model for user performance time. *Communication of the ACM* , 369-410.

Carroll, J. (2003). Introduction: toward a multidisciplinary science of human-computer interaction. In J. Carroll, *HCI models, theories, and frameworks* (pp. 1-11). Amsterdam: Elsevier Science.

Cassel, R. N., & Cassel, S. L. (1984). Cassel computer literacy test (CMLTRC). *Journal of Instructional Psychology , 11*, 3-9.

Caulton, D. A. (2001). Relaxing the homogeneity assumption in usability testing. *Behaviour & Information Technology , 20* (1), 1-7.

Chapanis, A. (1991). Evaluating usability. In B. Shackel, & S. Richardson, *Human factors in informatics usability* (pp. 360-395). Cambridge: Cambridge University Press.

Chen, C., Czerwinski, M., & Macredie, R. (2000). Individual differences in virtual environments - Introduction and overview. *Journal of American Society for Information Science* , 499-507.

Churchill, G. A. (1979). A Paradigm for Developing better Measures of Marketing Constructs. *Journal of Marketing Research , 16*, 64-73.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in scale development. *Psychological Assessment, . , 7*, 309-319.

Compeau, D. R., & Higgins, C. A. (1995). Computer self-efficacy: Development of a measure and initial test. MIS Quarterly,. *19* (2), 189-211.

Connell, I., Blanford, A., & Green, T. (2004). CASSM and cognitive walkthrough: usability issues with ticket vending machines. *Behaviour & Information Technology, , 23* (5), 307-320.

Cooper, A., & Reimann, R. (2003). *About face 2.0: The essentials of interaction design.* Indiana: Wiley.

Cooper, C. (1998). *Individual differences.* London: Arnold.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York : Holt.

Cronbach, L., & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin , 52*, 281-302.

Dawis, R. (1987). Scale Construction. Journal of Counseling Psychology. *34*, 481-489.

DeVellis, R. (1991). *Scale development: Theory and Application.* Newbury Park, CA: Sage.

Diamantopoulos, A., & Winklhofer, H. (2001). Index Construction with Formative Indicators: An Alternative to Scale Development. *Journal of Marketing Research , 38* (2), .269-277.

Dillon, A., & Watson, C. (1996). User analysis in HCI - the historical lessons from individual differences research. *International Journal of Human-Computer Studies* , 619-637.

Dumas, J. S., & Redish, J. C. (1993). *A practical guide to usability testing.* Ablex: Norwood- NJ.

Dunnette, M. (1976). *Handbook of industrial and organizational psychology.* Chicago: Rand McNally College Publishers.

Dwivedi, Y., Banita, L., Williams, M., Schneberger, S., & Wade, M. (2009). *Handbook of research on contemporary theoretical models in information systems.* London: Information Science Reference.

Egan, D. E. (1988). Individual differences in human-computer interaction. In M. Helander, *Handbook of human-computer interaction* (pp. 543-565,). New York : Elsevier, .

Egan, D., Bowers, C., & Gomez, L. (1982). Learner characteristics that predict success in using a text-editor tutorial. *Proceedings of the 1982 Conference on Human Factors in Computing Systems*, (pp. 337-340).

Evans, G., & Simkin, M. (1989). What best predicts computer proficiency. *Communications of the ACM* , 1322-1327.

Freudenthal, D. (2001). The role of age, foreknowledge and complexity in learning to operate a complex device. *Behavior & Information Technology, , 20* (1), 23-35.

Garmer, K., Liljegren, E., Osvalder, A., & Dahlman, S. (2002). Application of usability testing to the development of medical equipment. *International Journal of Industrial Ergonomics, , 29*, 145-159.

Gomez, L. M., Egan, D. E., Wheeler, E. A., Sharma, D. K., & Gruchacz, A. (1983). How interface design determines who has difficulty learning to use a text editor. *CHI' 83 Proceedings*, (pp. 176-181).

Gray, W., & Salzman, M. (1998). Damaged merchandize? A review of experiments that compare usability evaluation tools. *Human-Computer Interaction* , 203-261.

Haynes, S. N., Richard, D. C., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment , 7*, 238-247.

Helander, M. (1997). The human factors profession. In G. Salvendy, *Handbook of Human Factors and Ergonomics* (pp. 3–16,). New York: Wiley.

Hogan R. T. (1991). Personality & personality measurement. In D. M. D., & H. L. M., *Handbook of industrial and organizational psychology* (pp. 873–919.). Palo Alto, CA: Consulting Psychologists Press.

Igbaria, M., Zinatelli, N., Cragg, P., & Cavaye, A. L. (1997, September). Personal computing acceptance factors in small firms: a structural equation model. *MIS Quarterly* , 279-305.

International Standards Organization. (2006). ISO 20282 - Ease of operation of everyday products.

Johnson, J. A. (2004). The impact of item characteristics on item and scale validity. *Multivariate Behavioral Research , 39*, 273-302.

Jones, M., & Pearson, R. (1996). Developing an instrument to measure computer literacy. *Journal of Research on Computing Education* , 17-29.

Jöreskog, K., & Sörbom, D. (1993). *Lisrel 8: Structural equation modeling with the SIMPLIS command language.* Lincolnwood, IL: Scientific Software International.

Kay, R. H. (1993). An exploration of theoretical and practical foundations for assessing attitudes towards computers: the computer attitude measure (CAM). *Computers in human behavior , 19*, 11-56.

Kiesler, S., Zdaniuk, B., Lundmark, V., & Kraut, R. (2000). Troubles with the internet: The dynamics of help at home. *Human-Computer Interaction* , 323-351.

Kinzie, M. B., Delcourt, M. A., & Powers, S. M. (1994). Computer technologies:attitudes and self-efficach across undergraduate disciplines. *Res. Higher Education , 35* (6), 745-768.

Kleinbaum, D., & Kupper, L. (1978). *Applied regression analysis and other multivariable methods.* Massachusetts: Duxbury Press.

Kline, R. (2005). *Principles and practice of structural equation modeling.* New York: Guilford Press.

Kline, R. (2005). *Principles and practice of structural equation modeling.* New York: Guilford Press.

Koschmann, T., Kuuti, K., & Hickman, L. (1998). The concept of breakdown in Heidegger,Leont'ev, and Dewey and its implications for education. *Mind, Culture, and Activity , 5* (1), 25-41.

Krosnick, J. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology , 5*, 213-236.

Krosnick., J. A., Judd, C. M., & Wittenbrink, B. (2005). The measurement of attitudes. In D. Albarracín, B. T. Johnson, & M. P. Zanna, *The Handbook of Attitudes.* (pp. 21-76). Mahwah, NJ: Erlbaum.

Lansdale, M. W., & Ormerod, T. C. (1994). *Understanding interfaces: A handbook.*

Lauretta, D., & Deffner, G. (1996). Experimental evaluation of dialogue styles for hybrid telephone-based interfaces. *Behaviour & Information Technology, , 15* (1), 51-56.

Loyd, B. H., & Loyd, D. E. (1985). The reliability and validity of an instrument for the assessment of computer attitudes. *Educational and Psychological Measurement , 45*, 903-908.

MacCallum, R., & Browne, M. (1993). The use of causal indicators in covariance structure models: some practical issues. *Psychological Bulletin , 114* (3), 533-41.

Mack, R., & Montaniz, F. (1994). Observing, predicting, and analyzing usability problems . In Nielsen, *Usability inspection methods* (pp. 295 – 341).

Marakas, G. M., Yi, M. Y., & Johnson, R. D. (1998). The multilevel and multifaceted character of computer self-efficacy: Toward clarification of the construct and an integrative framework for research. *Information systems research , 9* (2), 126-162.

McDonald, R., & Moon-Ho, R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods* , 64-82.

Murphy, C. A., Coover, D., & Owen, S. V. (1989). Development and validation of the computer self-efficacy scale. *Educational and Psychological Measuremen , 49*, 893-899.

Netemeyer, R., Bearden, W., & Sharma, S. (2003). *Scaling procedures.* Newburry Park, CA: Sage.

Newell, A. &. (1972). *Human problem solving.* Englewood Cliffs: Prentice Hall.

Nielsen, J. (1994). Heuristic evaluation. In J. Nielsen, & R. L. Mack, *Usability Inspection Methods* (pp. 25-62). New York: John Wiley & Sons.

Nielsen, J. (1993). *Usability engineering. .* Boston: Academic Press.

Nilsen, E., Jong, H., Olson, J. D., Biolsi, K., Rueter, H., & Mutter, S. (1993). The growth of software skill: A longitudinal look at learning & performance. *Proceedings of Interchi' 93*, (pp. 149-156).

Norman, D. A. (1988). *The design of everyday things. .* New York: Doubleday.

Novick, D., & Ward, K. (2006). Why don't people read the manual? *Proceedings of the 24th annual ACM international conference on design of communication* (pp. 11-18). Myrtle Beach: ACM.

Nunnally, J. (1978). *Psychometric theory* (2nd edition ed.). New York: McGraw Hill.

Oskamp, S. (2004). *Attitudes and opinions. .* Mahawa, NJ: Erlbaum.

Pajares, F. (1997). Current directions in self-efficacy research. In M. Maehr, & P. R. Pintrich, *dvances in motivation and achievement.* (Vol. 10 , pp. 1-49.).

Peter, J. P. (1981). Construct validity: A review of basic issues and marketing practices. *Journal of Marketing Research , 18*, 133-145.

Potosnak, K. (1988). Recipe for a usability test. *IEEE Software* , 83-84.

Preece, J. (1994). *Human-computer interaction.* Harlow: Addison-Wesley.

Proctor, R. W., & Dutta, A. (1995). *Skill acquisition and human performance.* London: Sage.

Quade, A. (2003). Development and validation of a computer science self-efficacy scale for CS0 courses and the group analysis of CS0 student self-efficacy. *Proceedings of the international conference on information technology ITCC'03.*

Raykov, T., & Marcoulides, G. (2006). *A first course in structural equation modeling.* New Jersey: Erlabaum.

Rettig, M. (1991). Nobody reads documentation. *Communications of the ACM archive* , 19-24.

Ribak, R. (2001). 'Like immigrants': Negotiating power in the face of the home computer. *New Media Society* , 220-238.

Richardson, S., & Shackel, B. (1991). *Human factors for informatics usability.* Cambridge: Cambridge University.

Schumacker, R., & Lomax, R. (2004). *A beginner's guide to structural equation modelling.* New Jersey: Lawrence Erlbaum.

Shackel, B., & Richardson, S. (1991). *Human factors for informatics usability. .* Cambridge .

Smith. (1997). *Human-computer factors.* London: McGraw-Hill.

Smith, G., & McCarthy, D. (1995). Methodological considerations in the refinement of clinical assessment instruments. *Psychological Assessment , 7* (3), 300-308.

Spector, P. (1992). *Summated rating scale construction: An introduction.* Newbury Park,CA: Sage.

Spradley, J. (1979). *The ethnographic interview.* Fort Worth: Harcort Brace Jovanovich College Publishers.

Stat Notes. (n.d.). Retrieved February 2, 2010, from http://faculty.chass.ncsu.edu/garson/PA765/structur.htm

Sternberg, R. J. (1999). *Cognitive Psychology. .* Harcourt Brace College.

Sutcliffe, A., Ryan, M., Doubleday, A., & Springett, M. (2000). Model mismatch analysis: towards a deeper explanation of users' usability problems. . *Behaviour & Information Technology* , 19 (1), 42-55.

Şimşek, Ö. (2007). *Yapısal Eşitlik Modellemesine Giriş: Temel İlkeler ve LISREL Uygulamaları.* Ankara: Ekinoks.

Thimbleby, H. (. (1991, February). Can anyone work the video. *New Scientist* , 40-43.

Thompson, B. (2000). Ten commandments of structural equation modeling. In L. Grim, & P. Yarnold, *Reading and understanding multivariate statistics* (pp. 261-283). Washington, DC: American Psychology Association.

Torkzadeh, G., & Van Dyke, T. P. (2001). Development and validation of an internet self-efficacy scale. *Behaviour & Information Tehcnology , 20* (4), 275-280.

Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin , 103*, 299-314.

Uebersax, J. (2000). *Agreement on interval-level ratings.* Retrieved May 28, 2008, from http://ourworld.compuserve.com/homepages/jsuebersax/cont.htm

Vincente, K. J., Hayes, B. C., & Willigies, R. C. (1987). Assaying and isolating individual differences in searching a hierarchical file system. *Human Factors* , 349-359.

Vygotsky, L. S. (1978). *Mind in society: the development of higher psychological processes.* Cambridge: Harvard University Press, .

Woodworth, R. (1939). *Experimental Psychology.* London: Methuen.

Wu, L., & Rocheleau, B. (2001). Formal versus informal end user training in public and private sector organizations. *Public Performance & Management Review* , 312-321.

Zhang, Y. (2006). *Unstructured interview.* Retrieved February 23, 2009, from www.ils.unc.edu/~yanz/Unstructured%20interview.pdf

APPENDIX A

LEARNING ELECTRONIC DEVICES QUESTIONNAIRE SAMPLE FORM

Cinsiyet: __ Kadın - __ Erkek
Yaş:
Eğitim durumu: __ İlkokul - __ Ortaokul - __ Lise - __ Üniversite - __ Y. Lisans
Meslek:

Katılacağınız bu araştırma Orta Doğu Teknik Üniversitesi, Endüstri Ürünleri Tasarımı Bölümü'nde devam etmekte olan bir doktora çalışması için veri toplamayı amaçlamaktadır. Araştırma ticari bir amaç gütmemektedir. Elde edilecek sonuçlar tamamen akademik amaçlar doğrultusunda değerlendirilecektir.

Elektronik Aletlerin Öğrenimi

Elektronik aletlerin yaşantımızdaki yeri günden güne artmaktadır. Elektronik aletleri artık hem günlük işlerimizde, gereksinimlerimizin karşılanmasında, hem de eğlence amacıyla daha çok kullanıyoruz. Bu durumun sonucu olarak her geçen gün daha fazla ürünün nasıl kullanıldığını öğrenmek zorunda kalıyoruz.

Araştırmamızın amaçlarından biri kullanıcıların elektronik aletleri öğrenimlerini kolaylaştıran ve zorlaştıran etkenleri saptamaktır. Çalışma kapsamında 'elektronik alet' olarak tanımlanan ürünler dijital ekranları veya elektronik panelleri bulunan, menüler aracılığıyla kullanılan dijital ürünlerdir.

'Elektronik aletler' grubuna giren ürünlere aşağıdaki örnekler verilebilir:

Cep telefonları;
Dijital fotoğraf makinaları, video kameralar;
Dijital ekranlı fırın, çamaşır makinası, bulaşık makinası gibi ürünler;
MP3 çalarlar;
Elektronik ajandalar, sözlükler;
DVD/VCD oynatıcılar ve kaydediciler, uydu alıcılar;
Müzik setleri;
Televizyonlar;
Elektronik oyunlar;
Bankamatik cihazları

İkinci sayfada yer alan sorular doğrultusunda bu konudaki fikirlerinizi aktarırsanız yapılan araştırmaya önemli bir katkıda bulunmuş olacaksınız.

>>> Lütfen bir sonraki sayfaya geçiniz >>>

1) Lütfen ne gibi durumlarda elektronik aletlerin **kullanımını daha kolay öğrendiğinizi** belirtiniz...

Örnek:
Kullanım kılavuzu açık yazılmışsa önce kılavuzu okurum ve aleti kolayca çözerim.
Alet çok fazla özelliğe sahip değilse çabuk öğrenirim.

2) Lütfen ne gibi durumlarda elektronik aletlerin **kullanımını daha zor öğrendiğinizi** belirtiniz...

Örnek:
Takıldığımda yardım edecek kimse yoksa daha zor öğrenirim.
Üründen çok hoşlanmamışsam öğrenmekte zorlanırım.

Çalışmamıza katıldığınız için çok teşekkür ederiz...

Eğer bu formu elektronik ortamda cevapladıysanız lütfen dosyayı ali.berkman@gmail.com adresine epostayla iletiniz...

APPENDIX B

POSITIVE AND NEGATIVE EXPRESSIONS COMPILED AFTER LEDQ

Novelty – familiarity > familiar product family

	Effect	Expressions	f*[62]
1	**+**	“Daha önce kullandığım tür aletse”	1
2	**-**	“Daha önceden kullanmadığım bir tür aletse”	1
3	**+**	“Aynı işe yarayan bir alet kullandıysam”	1
4	**-**	“Daha önce karşılaşmadığım bir ürünse”	1

Novelty – familiarity > familiar interface / product

1	**+**	“Bildiğim bir aletin sistemiyle aynıysa”	1
2	**+**	“Daha önceden kullandığım aletlere benziyorsa”	8
3	**+**	“Daha önce kullandığım aletlerin kullanımına benziyorsa”	1
4	**+**	“Sık sık kullandığım bir alete benziyorsa”	1

[62] number of times the argument is expressed

5	**+**	"Diğer aletlerden bildiğim kullanım mantığını uygulayabiliyorsam"	1
6	**+**	"Çok değişik özelliklere sahip değilse"	1
7	**+**	"Menüsü benzer ürünlerle paralel yapıdaysa"	1
8	**-**	"Diğer ürünlerle benzerlik taşımıyorsa"	1
9	**+**	"Önceki tecrübemi kullanabiliyorsam"	1
10	**-**	"Standart dışı tasarımı olan bir ürünse"	1
11	**-**	"Farklı kullanılan tuşları, kontrolleri varsa"	1
12	**-**	"Çok farklı bir aletse"	1
13	**-**	"Modern bir aletse"	1
14	**-**	"Tuşlar genelde kullanılan amaçlara tersse"	1
15	**+**	"Daha önce benzer bir menüyle karşılaşmışsam"	1
16	**-**	"Daha önce kullandığım aletlerden çok farklıysa"	2
17	**-**	"Bana yabancı bir ürünse"	1

Novelty – familiarity > familiar brand

1	**+**	"Alıştığım bir markanın ürünüyse"	1
2	**+**	"Aynı markanın başka ürünlerini kullanmışsam"	1

3	**-**	"Yepyeni bir markaysa"	1
4	**+**	"Herkes tarafından tercih edilen bir markaya aitse"	1
5	**-**	"Bilinen, tanınan bir marka değilse"	1
6	**+**	"Piyasada en çok satılan markaysa"	1

Novelty – familiarity > similarity with previous model

1	**+**	"Mevcut olan bir modelin yeni versiyonuysa"	1
2	**+**	"Daha önceki modelleriyle benzerlik gösteriyorsa"	1
3	**+**	"Eski modelin üstüne eklemeler yapılmışsa"	1
4	**-**	"Eski aleti değiştirip yeni bir alet aldığım zaman"	1
5	**-**	"Daha önceden farklı bir model kullanmışsam"	1
6	**-**	"Daha önce alıştığım aletle arasında çok fark varsa"	1
7	**-**	"Önce kullandığım modelden farklı görünüyorsa"	1

Novelty – familiarity > diffusion

1	**+**	"Çok kişi tarafından kullanıldığı için göz aşinalığı oluştuysa"	1

2	**-**	“Aletin kullanımı yaygın değilse”	1
3	**-**	“Yeni teknlojiler içeriyorsa”	1
4	**-**	“Çok yeni bir aletse”	3
5	**-**	“Aletin ilk kullanıcılarındansam”	1
6	**-**	“Yaygın olmayan bir ürünse”	1
7	**+**	“Genellikle çoğunluk tarafından biliniyorsa”	1
8	**-**	“Kullanımı yaygın bir ürün değilse”	1

Affection > interest

1	**-**	“İlgimi çekmemişse”	4
2	**-**	“İlgi çekici gelmediğinde”	2
3	**-**	“Çok ilgilenmediğim bir aletse”	1
4	**-**	“İlgi alanıma girmiyorsa”	8
5	**+**	“İlgi alanıma giriyorsa”	4
6	**+**	“Alete karşı ilgim fazlaysa”	1
7	**-**	“İlgim azaldıysa”	1

8	**-**	"Ürüne ilgi duymuyorsam"	1

Affection > emotion

1	**+**	"Sevdiğim bir ürünse"	1
2	**+**	"Hoşlandığım bir ürünse"	1
3	**-**	"Üründen çok hoşlanmadığım zamanlarda"	1
4	**-**	"Ürüne çok fazla ısınamadıysam"	1
5	**-**	"Ürünü çok fazla sevmediysem"	1
6	**-**	"Üründen çok hoşlanmamışsam"	4
7	**-**	"Alete karşı tepkiliysem"	1
8	**-**	"Öğrenme isteksizliği varsa"	2
9	**+**	"Kullanmayı gerçekten istiyorsam"	1
10	**+**	"Öğrenmeyi gerçekten istiyorsam"	1
11	**-**	"Öğrenme isteğim çok değilse"	1
12	**-**	"Öğrenmekten zevk almıyorsam"	1
13	**+**	"Nasıl kullanıldığını çözmek hoşuma gidiyorsa"	1
14	**-**	"Kullanmak istemiyorsam"	1

15	**-**	“Ürünü kullanmak beni sıkıyorsa”	1
16	**-**	“Öğrenmekten çabuk sıkılıyorsam”	1
17	**+**	“Alet bende merak uyandırıyorsa”	1
18	**-**	“Alet bana itici geliyorsa”	1
19	**+**	“Kullanıcıya hitabeden bir aletse”	1
20	**-**	“Severek aldığım bir ürün değilse”	1

Affection > visual appeal

1	**-**	“Görsel açıdan keyif vermeyen bir ürünse”	1
2	**-**	“Rengi çekici değilse”	1
3	**-**	“Aletin görünüşünü sevmemişsem”	1
4	**+**	“Güzel tasarlanmış bir ürünse”	1
5	**+**	“İlginç bir görünümü varsa”	1

Usefulness > need

1	**-**	“Çok gerek görmediğim bir aletse”	1

2	**-**	“Ürünün özelliklerini çok fazla kullanmayacaksam”	1
3	**+**	“Kişisel isteklerime uygun özellikleri varsa”	1
4	**-**	“Ürüne fazla ihtiyaç duymuyorsam”	1
5	**+**	“İhtiyaçlarımı karşılayacaksa”	1
6	**+**	“İhtiyaçlarıma cevap verecek nitelikteyse”	1
7	**+**	“Alet ihtiyaçtan alınmışsa”	1
8	**+**	“Günlük yaşantımı kolaylaştıracak nitelikteyse”	1
9	**+**	“İhtiyaçlarıma cevap vermiyorsa”	2
10	**-**	“İhtiyaçtan ötürü edinmemişsem”	1
11	**-**	“Günlük hayatta kullanabileceğim bir şey değilse”	1
12	**+**	“Kullanmayacağım fonksiyonları yoksa”	1
13	**-**	“İşime yaramayacak fonksiyonları, özellikleri çoksa”	1
14	**-**	“İşime yaramayacak bir ürünse”	2
15	**+**	“İşime yarıyorsa “	1
16	**+**	“Aletin ilgilendiğim kısımları çoksa”	1
17	**+**	“İşlevselliği iyiyse”	1
18	**-**	“İşlevselliği iyi değilse”	1

19	**+**	“İşimi daha iyi yapmam için gerekli bir aletse”	*
20	**+**	“Yaptığım işleri daha iyi yapmamı sağlayacaksa”	*

* not directly expressed by respondents

Usefulness > neccesity

1	**-**	“Günlük hayatta sürekli kullanmayacağım bir aletse”	1
2	**-**	“Kullanmak zorunda olmadığım bir ürünse”	1
3	**-**	“Kullanımı çok elzem değilse”	1
4	**+**	“Günlük hayatta çok kullandığım bir aletse”	2
5	**+**	“Aleti kullanmam gerekiyorsa”	1
6	**+**	“Yaşantımı çok etkileyecek bir aletse”	1
7	**+**	“Sıkça kullandığım bir ürünse”	2
8	**-**	“Sürekli kullanmam gerekmiyorsa”	1
9	**-**	“Kullanmak zorunda bırakıldıysam” [was previously listed under urgency]	1

Usefulness > urgency

1	**+**	"Aleti kullanmaya mecbursam"	1
2	**+**	"Aleti kullanmaktan başka çarem yoksa"	1
3	**+**	"Çok acelem olduğu zamanlarda"	1
5	**-**	"Hızlı bir şekilde öğrenmem gerekiyorsa"	1
6	**+**	"Acilen öğrenmem gerekiyorsa"	1
7	**+**	"Çok zor durumdaysam"	1

Ease of use [general]

1	**+**	"Basit bir tasarıma sahipse"	1
2	**+**	"Tasarımı iyiyse"	1
3	**-**	"Tasarımı kötüyse"	1
4	**+**	"Basit bir aletse"	1
5	**+**	"Yanlış yaptığımda uyarı gelirse"	1
6	**+**	"Ekranından yazıyla bilgi veriliyorsa"	1
7	**+**	"Menü mantığı bana ters gelmiyorsa"	1
8	**+**	"Menü mantığı basitse"	1
9	**+**	"Yanlış yaptığımda uyarı gelirse"	1

10	**-**	"Mantığı sağlam değilse"	1
11	**+**	"Menüler anlaşılırsa"	1
12	**+**	"Menüsü açıksa"	2
13	**+**	"Kolay okunabilir bir menüye sahipse"	1
14	**+**	"Basit bir arayüzü varsa"	2
15	**+**	"Kullanımı pratikse"	1
16	**+**	"Menü kullanımı kolaysa"	1
17	**+**	"Nasıl kullanılacağı açıksa"	1
18	**+**	"Kolay kullanılabilen bir aletse"	5
19	**+**	"Basit tasarlanmışsa"	1
20	**+**	"Basit adımlarla istediğime ulaşabiliyorsam"	1
21	**+**	"İlk görüşte basit olduğuna inandıysam"	1
22	**-**	"Kullanım açık değilse"	1
23	**-**	"Nasıl kullanılacağı net değilse"	1
24	**-**	"Özellikleri kolayca kullanılamıyorsa"	1
25	**-**	"Kullanımı zor bir aletse"	1
26	**-**	"Kullanışsız bir ürünse"	1
27	**-**	"Aletin kullanımı karışıksa	4

28	**-**	“Menü kullanımı zorsa”	1
29	**-**	“Pratik değilse”	1
30	**-**	“Basit tasarlanmamışsa””	1
31	**-**	“Arayüzü anlaşılmazsa”	1
32	**+**	“Çok kullanılan fonksiyonlar kolay bulunuyorsa”	1
33	**+**	“Kullanım aşamaları akılda kalıcıysa“	1
34	**+**	“Menülerde her işlemin düzgün sırayla yerleştirilmiş olması”	1
35	**-**	“Ürünün çalışma biçimini kavrayamadıysam”	1
36	**-**	“Tuşların fonksiyonlarını kavrayamadıysam”	1

Ease of use> efficiency

1	**+**	“Kısa yolları varsa”	1
2	**+**	“Kısa yolları yoksa”	1
3	**+**	“Sonuca kolay götürecek menüsü varsa”	1
4	**+**	“İşlemler tek tuşla yapılabiliyorsa”	1
5	**+**	“Hızlı bir şekilde istediğime ulaşabiliyorsam”	1
6	**-**	“Kullanım dolambaçlı olursa”	1

7	**-**	“Kullanım sırasında bir sürü aşamadan geçmek gerekiyorsa”	1
8	**+**	“Özelliklere hemen ulaşabiliyorsam”	1

Ease of use> intuitiveness

1	**-**	“Tuşların açıklamaları yoksa”	1
2	**+**	“Tuşların üstünde ne işe yaradıkları yazıyorsa”	1
3	**-**	“Kullanılan semboller belirgin değilse”	1
4	**-**	“Tuşların üstündeki açıklamalar diğer aletlerden farklıysa”	1
5	**-**	“Sık sık kılavuza başvurmam gerekiyorsa”	1
6	**-**	“İç güdülerime dayanarak çözemiyorsam”	1
7	**+**	“Kullanım sırasında düzgün yönlendirmeler yapılıyorsa”	1
8	**+**	“Menülerde direktifler açıksa”	1
9	**+**	“Menülerde açıklayıcı bilgiler varsa”	1
10	**+**	“Menüde ikonlar (küçük resimler) kullanıldıysa”	1
11	**+**	“Basitçe mantık yürüterek çözebileceğim bir aletse”	1
12	**+**	“İlk bakışta nasıl kullanılacağını anlıyorsam”	1
13	**+**	“Aletin üsünde ikonlar bulunuyorsa“	1

14	**+**	“Temel fonksiyonlar aletin üstünde belirgin şekilde gösterilmişse”	1
15	**+**	“Aletin üstünde işaretler bulunuyorsa”	1
16	**+**	“Simgelerden çalışma mantığını anlayabiliyorsam”	1
17	**+**	“Kılavuza ihtiyaç duymadan alet kendi kendini anlatabiliyorsa”	1
18	**+**	“Kullanılan ikonlar anlatılmak istenen konuyu çağrıştırıyorsa”	1
19	**-**	“Anlaşılmayan semboller olursa”	1
20	**-**	“Tuşların ne işe yaradığı anlaşılmıyorsa”	1
21	**-**	“Menü üzerindeki işaretler tanıdık olmazsa”	1
22	**-**	“Menü üzerindeki harfler tanıdık olmazsa”	1
23	**-**	“Aletin üstünde belirsiz açıklamalar olursa”	1
24	**+**	“Kullanım şekli ön yüzde gösteriliyorsa”	1
25	**+**	“Aletin üzerindeki yazılar açıklayıcıysa”	1
26	**-**	“Aletin üzerindeki yazılar yönlendirici değilse”	1
27	**+**	“Aletin üstünde yönlendirici bilgiler olursa”	1
28	**+**	“Kullanım sırasında uygun yönlendirici bilgiler verilirse”	1

Ease of use> physical characteristics

1	**+**	"Tek bir düğmesi varsa"	1
2	**+**	"Tek tuşla kullanılabiliyorsa"	2
3	**-**	"Tuşların birden fazla işlevi varsa"	1
4	**-**	"Çok fazla düğmesi varsa"	2
5	**-**	"Çok fazla tuşu varsa"	2
6	**+**	"Fazla tuşu yoksa"	1
7	**+**	"Geniş bir ekranı varsa"	1
8	**+**	"Fonksiyonlar net bir şekilde düğmelerle tanımlanmışsa"	1
9	**+**	"Belirli fonksiyonlar için belirli tuşlar varsa"	1
10	**+**	"Çok fazla tuşu yoksa"	1
11	**-**	"Kullanım paneli ürünün görünmeyen yerlerindeyse"	1
12	**-**	"Ön panel karmaşık görünümlüyse"	1
13	**+**	"İlgili düğmeler birbirine yakın yerleştirilmişse"	1
14	**-**	"Tuşlar çok küçük olduğu için rahat kullanamıyorsam"	1
15	**-**	"Yazılar ve rakamlar büyük değilse"	1

Ease of use> simplicity >structure

1	**+**	“Menüsü çok karışık değilse”	3
2	**+**	“Alet karmaşık bir yapıya sahip değilse”	1
3	**+**	“Zincirleme olarak alt menülere girilmesi gerekmiyorsa”	1
4	**+**	“Fazla karmaşık değilse”	2
5	**+**	“Az detay içeriyorsa”	1
6	**+**	“Çok komplike değilse”	2
7	**-**	“Menülerde çok fazla değişken varsa”	1
8	**-**	“Menüsü çok karışıksa”	5
9	**-**	“Alette çok menü varsa”	1
10	**-**	“Fazla alt menüsü olduğu için sıkılırsam”	1
11	**-**	“Menüsü sürekli alt açılımlar veriyorsa”	1
12	**-**	“Menüler çok fazla karışık yapılmışsa”	1
13	**-**	“Menülerin içeriği çok fazlaysa”	1
14	**-**	“Menüler çok karmaşık olursa”	4
15	**-**	“Çok detaylıysa”	1
16	**-**	“Alet çok karmaşık özelliklere sahipse”	1
17	**-**	“Çok ayrıntılı özelliklere sahip olması”	1
18	**-**	“Alet karmaşıksa”	4

19	**-**	“Çok komplike bir aletse”	3
20	**-**	“Kompleks bir aletse”	2
21	**+**	“Fonksiyonel yapı iyi basamaklandırılmışsa”	1
22	**+**	“Özellikler iyi yerleştirilmişse”	1
23	**+**	“Menülerin içeriği azsa”	1
24	**-**	“Karmaşık görünüyorsa”	1

Ease of use> simplicity >number of functions

1	**+**	“Fazla özelliğe sahip değilse”	2
2	**+**	“Çok fazla özelliğe sahip değilse”	3
3	**+**	“Az özelliği varsa”	1
4	**+**	“Alet az fonksiyonluysa”	1
5	**-**	“Çok fonksiyonluysa”	2
6	**-**	“Çok amaçlı bir ürünse	1
7	**-**	“Çok fazla özelliğe sahipse”	4
8	**-**	“Eğer çok programlıysa”	1

Ease of use> language >literal

1	**+**	"Özellikler, fonksiyonlar iyi adlandırılmışsa"	1
2	**+**	"Özellikler iyi adlandırılmamışsa"	1
3	**-**	"Kullanılan teknik kelimeler anlaşılmaz olursa"	1
4	**-**	"Üst menülerle alt menülerin isimleri uyumlu değilse"	1
5	**-**	"Menü başlıklarını anlamlı değilse"	1
6	**-**	"Ürünün üstünde anlaşılmayan günlük hayatta kullanılmayan sözcükler varsa"	1
7	**+**	"Ürün kullanıcının dilinden konuşuyorsa "	1
8	**+**	"Menülerde dil seçeneği varsa"	1
9	**+**	"Türkçe menülüyse"	2
10	**+**	"Tuşların üstünde Türkçe yazılar varsa"	1
11	**-**	"Üründe bilmediğim bir dil kullanılıyorsa"	2
12	**-**	"Üründe dil karmaşası varsa"	1
13	**+**	"Alette kullanılan dil açıksa"	1
14	**-**	"Dil düzgün değilse"	1

Ease of use> language >literal

1	**+**	“Menülerde şekiller kullanılmışsa”	1
2	**+**	“Menülerde resimler kullanılmışsa”	1
3	**+**	“Menüleri renkliyse”	1
4	**-**	“Menülerde dikkat çekici unsurlar varsa”	1
5	**-**	“Menülerde düz siyah yazılar kullanılmışsa”	1

Help and support > informal help > from salespeople

1	**+**	“Satın aldığım yerden kullanım önerileri alabilirsem”	1
2	**+**	“Satın aldığım yerde öğreten biri varsa”	1
3	**+**	“Satılırken açıklayıcı bilgi verilirse”	2
4	**-**	“Satan yer yardımcı olmazsa”	1
5	**+**	“Satıcı nasıl kullanacağımı gösterirse”	1
6	**+**	“Satış elemanı yardımcı oluyorsa”	1

Help and support > informal help > user forums

1	**+**	“Aletle ilgili forumlar varsa”	1

Help and support > informal help > to others

1	**-**	“Ürünü öğrenip başkasına öğretmek zorundaysam”	1
2	**-**	“Ürünü başkası için kullanmam gerekiyorsa”	1
3	**-**	“Ürünü çabuk kurmam ve kullanmam isteniyorsa”	1

Help and support > informal help > from others

1	**+**	“Aleti kullananlardan bilgi alabilirsem”	3
2	**+**	“Bilen kişilere sorabiliyorsam”	1
3	**+**	“Bilen biri tarafından kullanım anlatılırsa”	3
4	**+**	“Biri bana nasıl kullanıldığını özetleyebilirse”	1
5	**+**	“Bilen biri gösterdiğinde”	3
6	**+**	“Ürünü daha önce kullanmış bir arkadaşım varsa”	1

7	**+**	"Zorlandığımda yardım alabileceğim biri olursa"	1
8	**+**	"Kullanabilen birini gözlemleme şansım varsa"	1
9	**+**	"Merakı olan birinden destek alabiliyorsam"	1
10	**+**	"Tanıdığım biri aleti bana öğretirse"	1
11	**+**	"Bilen birinden yardım alabilirsem"	1
12	**-**	"Öğrenmemi destekleyecek biri yoksa"	1
13	**+**	"Daha önce kullananlardan destek alırsam"	1
14	**+**	"Daha önce kullananlara danışma fırsatım varsa"	1
15	**+**	"Kullanımı bilen bir uygulamalı olarak anlatırsa"	1
16	**+**	"Kullanan biri anlatırsa"	1
17	**+**	"Uzman bir kişi anlatırsa"	1
18	**-**	"Yardım alabileceğim kimse yoksa"	3
19	**-**	"Kullanan başka insanlar yoksa"	1
20	**-**	"Takıldığım zaman yardım edecek kimse yoksa"	1
21	**-**	"Kullanımı gösterecek kişiler yoksa"	1
22	**-**	"Bilgi alabileceğim kimse yoksa"	1
23	**-**	"Bilen biri yoksa"	1

24	**-**	“Yönlendirecek biri olmadığında”	1
25	**-**	“Detaylı şekilde anlatacak biri yoksa”	1
26	**-**	“Anlatacak bir kişi yoksa”	1

Help and support > formal help > instruction manual >availability

1	**+**	“Kılavuzu varsa”	1
2	**+**	“Kılavuz yardımıyla “	1
3	**-**	“Kullanım kılavuzu yoksa”	3
4	**-**	“Herhangi bir kaynağa sahip değilsem”	1
5	**+**	“Rehberinden yardım alabiliyorsam”	1
6	**+**	“İyi bir yardım menüsüne sahipse”	1
7	**+**	“Kılavuzda 'hızlı başlangıç' gibi kısaca kullanımı anlatan bir bölüm varsa”	1
8	**+**	“Alet içinde kullanımı öğreten bir bölüm olursa”	1
9	**+**	“Kullanımı anlatan CD olursa”	1

Help and support > formal help > instruction manual > characteristics

1	**+**	“Kılavuz sade olursa”	1
2	**-**	“Kılavuz belirsiz olursa”	1
3	**-**	“Kılavuz anlaşılır değilse”	5
4	**-**	“Kılavuzda verilen bilgiler net değilse”	1
5	**-**	“Kılavuz iyi değilse”	1
6	**-**	“Kafa karıştırıcı bir kılavuzu varsa”	1
7	**-**	“Kılavuz üstünkörü hazırlanmışsa”	1
8	**-**	“Kılavuz yetersizse”	5
9	**-**	“Kullanım kılavuzu uzun anlatımlarla hazırlandıysa”	2
10	**-**	“Kılavuzda uzun sayfalar dolusu açıklamalara yer verildiyse”	1
11	**-**	“Kılavuz açık değilse”	4
12	**-**	“Kılavuz yeterince açıklayıcı değilse”	2
13	**-**	“Kılavuzda şemalarla anlatılmamışsa”	2
14	**-**	“Kılavuz fazla detaylıysa”	1
15	**-**	“Anlatım tarzı kötüyse”	1
16	**-**	“Kılavuzda gerekli bilgiler yoksa”	1

17	**-**	“Ürünün özellikleri kılavuzda açık anlatılmamışsa	1
18	**-**	“Herhangi bir kaynağa sahip değilsem”	1
19	**-**	“Kılavuzla ürün modeli uyuşmuyorsa”	1
20	**-**	“Kılavuzla kullanım birbirini tutmazsa”	1
21	**-**	“Kılavuzda kullanım adım adım tariflenmemişse”	1
22	**-**	“Ne yapmam gerektiği açık bir şekilde ifade edilmemişse”	1
23	**+**	“Kılavuzda yazanları tek tek uygulayabiliyorsam”	1
24	**+**	“Kılavuz ne yapılması gerektiğini kısaca anlatıyorsa”	2
25	**+**	“Kılavuzda kullanım neden sonuç ilişkisiyle anlatılıyorsa”	1
26	**-**	“Kılavuzda bilgiler neden sonuç ilişkisiyle anlatılmıyorsa”	1
27	**+**	“Kılavuzda basit talimatlar veriliyorsa”	1
28	**+**	“Kılavuz adım adım anlatıyorsa”	1
29	**+**	Kılavuz neyapılması gerektiğini tek tek ifade ediyorsa “	1
30	**+**	Kullanım kılavuzu kullanışlıysa”	1
31	**+**	Kullanım kılavuzu yeterince anlaşılabiliyorsa”	1
32	**+**	Kullanım kılavuzu açıklayıcıysa”	1
33	**+**	Kullanım kılavuzunu anlayabiliyorsam”	2

34	**+**	Kılavuz yeterince detaylıysa	2
35	**+**	Kılavuzda anlatılanlar üründe rahatça görülüyorsa”	1
36	**+**	Kılavuzda sadece aldığım ürün anlatılıyorsa”	1
37	**+**	İyi bir kullanım kılavuzuna sahipse”	1
38	**+**	Kılavuzda pratik kullanım bilgileri veriliyorsa”	1
39	**+**	Kılavuzda açıklamalar iyi yapılmışsa”	1
40	**+**	“Kılavuz anlaşılır olursa”	6
41	**+**	“Kılavuz net olursa”	2
42	**+**	“Kullanım kılavuzunda yalın bir dil kullanılmışsa”	1
43	**+**	“Kullanım kılavuzu açık olursa”	12
44	**+**	“Kullanım kılavuzu iyi düzenlenmişse”	1
45	**+**	“Kullanımı kolay bir kılavuzu olursa”	1
46	**+**	“Kullanma kılavuzunda çok basit şekilde anlatılmışsa”	2
47	**+**	“Kullanma kılavuzunda çok açık anlatılmışsa”	3
48	**+**	“Kullanım kılavuzu basit tablolarla anlatıyorsa”	1
49	**+**	“Kılavuz şekillerle anlatıyorsa”	5
50	**+**	“Kılavuzda şemalar olursa”	3

51	**+**	“Kılavuzda resimler olursa”	1
52	**-**	“Kullanım kılavuzunda günlük dilde kullanılmayan sözcükler bulunuyorsa”	1
53	**-**	“Kılavuz bilmediğim bir dilde yazılmışsa”	1
54	**-**	“Kullanım kılavuzu yabancı dille hazırlanmışsa”	1
55	**-**	“Kullanım kılavuzu İngilizce hazırlanmışsa*”	1
56	**+**	“Türkçe açıklamaları varsa”	1
57	**-**	“Kılavuz Türkçe olmazsa”	2
58	**+**	“Türkçe tercümesi başarılıysa”	1
59	**+**	“Düzgün bir Türkçe'yle çevrilmişse”	1
60	**-**	“Kılavuz yabancı dille yazılmışsa”	4
61	**+**	“Kılavuz Türkçe’yse”	2
62	**+**	“Kılavuzda kullanlılan dil açıksa”	1
63	**+**	“Kılavuzda kullanılan dil basitse”	1
64	**-**	“Kılavuzda teknik terimler kullanılıyorsa”	1
65	**+**	“Kılavuzda anlaşılır bir Türkçe kullanıldıysa”	1
66	**-**	“Kılavuzdaki dil kullanımı kötüyse”	2

Help and support > formal help > instruction manual >support services

1	**-**	“Internet sayfası yoksa”	1
2	**+**	“Internet sayfası varsa”	1
3	**+**	“Teknik servisten telefonla yardım alabiliyorsam”	1
4	**+**	“Teknik servise ulaşabiliyorsam”	1
5	**+**	“Müşteri hizmetlerini arayabiliyorsam”	1
6	**-**	“Teknik destek sistemi yoksa”	1
7	**-**	“Yardım merkezi yoksa”	1
8	**+**	“Danışma merkezi olursa”	1

Learning context and process >method

1	**+**	“Kılavuzu okursam”	5
2	**-**	“Kılavuzu hiç okuyamadıysam”	2
3	**+**	“Uygulama yapabiliyorsam”	1
4	**+**	“Deneme yanılma yöntemi uygulayabiliyorsam”	3
5	**-**	“İç güdülerimle dayanarak çözemiyorsam”	1

6	**-**	“Kılavuzdan okumadan öğrenmeye çalıştığımda”	1
7	**-**	“Deneme yanılmayla öğrenme şansım yoksa”	1
8	**-**	“Teorik anlatımlarla öğrenmek zorundaysam”	1
9	**-**	“Aletin kendisini görmeden öğrenmek zorundaysam”	1
10	**-**	“Denemeden sadece kullanımı anlatılarak öğrenmek zorunda kalırsam”	1
11	**-**	“Herşeyi tek tek denemek zorunda kalıyorsam”	1
12	**-**	“Kullanabilmek önce sayfalarca kılavuz okumam gerekiyorsa”	2

Learning context and process >achievement

1	**-**	“Bir kaç kullandığımda hala sorun yaşıyorsam”	1
2	**-**	“İlk kullanımda sorun yaşarsam”	1
3	**-**	“Eğer aletle ilgili bir sorun yaşadığım için tekrar yaşamaktan korkarsam”	1
4	**-**	“Kullanırken çok hata yapıyorsam”	1
5	**+**	“Çözmeye başladığımı hissedersem”	1

Learning context and process >opportunities

1	**-**	“Alete az zaman ayırabiliyorsam”	1
2	**-**	“Yeteri kadar uğraşma fırsatı bulamıyorsam”	1
3	**+**	“Öğrenmek için vaktim bolsa”	1
4	**-**	“Öğrenmek için zamanım çok darsa”	1
5	**+**	“Aleti sıkça kullanma fırsatı bulabiliyorsam”	1
6	**+**	“Aleti kurmak ve kaldırmak için uğraşmak gerekmiyorsa”	1
7	**-**	“Şarjı çok uzun gitmiyorsa”	1

Learning context and process >other users

1	**-**	“Öğrenmeye çalışırken yanımda bana müdahale eden biri olursa”	1
2	**-**	“Yanımda öğrenme konusunda benden daha becerikli biri varsa”	1
3	**-**	“Yanımda öğrenme konusunda benden daha hızlı bir varsa”	1
4	**+**	“Başkaları yanımdayken önce ben çözüyorsam”	1
5	**-**	“Yanımda zaten o aleti kullanmayı üstlenmiş biri varsa”	1

6	**-**	"Ürünü çabuk kurmam ve kullanmam isteniyorsa"	1
7	**+**	"Daha önce başkası tarafından kullanılmışsa"	1
8	**+**	"Daha önce başkası tarafından alınmışsa"	1
9	**-**	"Aletin karışık olduğunu daha önce birinden duyduysam"	1

Breakdowns>cost

1	**-**	"Alet pahalı olduğu için fazla deneme yapamazsam"	1
2	**-**	"Pahalı olduğu için deneme yanılma yöntemini kullanamıyorsam"	1
3	**-**	"Aletin bozulma riski yüksekse"	1
4	**-**	"Bozulabileceğini düşünürsem"	1
5	**-**	"Hemen bozulursa"	1
6	**-**	"Bozulmaya açık bir aletse"	1
7	**-**	"Bozulduğunda yaptırmak zorsa"	1
8	**-**	"Yanlış yaptığımda geri dönüş yoksa"	1
9	**-**	"Yanlış kullanıldığında başa dönmek zorsa"	1

Breakdowns>likelihood

1	**-**	"Çabuk arızalanacak bir alet olduğunu düşünüyorsam"	1
2	**-**	"Yanlış kullanıldığında arıza verirse"	1
3	**-**	"Hassas bir aletse"	1
4	**-**	"Kullanmaya çekindiğim bir aletse"	1
5	**-**	"Kullanmaktan korkuyorsam"	1
6	**-**	"Yanlış kullanıldığında başa dönmek zorsa"	1

Prior knowledge>terminology

1	**+**	"Kısaltmaların ne anlama geldiğini bilirsem"	1
2	**+**	"Terimlerin ne anlama geldiğini bilirsem"	1
3	**-**	"Çok fazla özel terim kullanılıyorsa"	1
4	**-**	"Çok fazla kısaltma kullanılıyorsa"	1

Prior knowledge>domain knowledge

1	**-**	“Gerekli bilgiye sahip değilsem”	1
2	**-**	“Gerekli alt yapım yoksa”	1
3	**-**	“Bilgi seviyeme uygun değilse”	1
4	**-**	“Daha önceden alet hakkında bilgim yoksa””	1
5	**-**	“Alet bilgi birikimim dışında bilgi gerektiriyorsa”	1
6	**-**	“Çok karışık bilgi içeriyorsa”	1

APPENDIX C

Positive and Negative Expressions Compiled after LEDQ (English)

WARNING: The expressions listed below were not translated using a systematic procedure and no data was collected in order to provide an English version of GISE-S. Therefore, following item stems should not be used for item generation or data collection.

Novelty – familiarity > familiar product family

	Effect	Expressions	f*[63]
1	**+**	"If it is a type of device that I used before"	1
2	**-**	"If it is a type of device that I didn't use before"	1
3	**+**	"If I used a device for a similar task"	1
4	**-**	"If it is a product that I didn't come across"	1

Novelty – familiarity > familiar interface / product

1	**+**	"If it has a similar system with a device that I know"	1
2	**+**	"If it resembles devices that I used before"	8
3	**+**	"If its use is similar to devices that I used before"	1
4	**+**	"If it is similar to a device that I often use"	1

[63] number of times the argument is expressed

5	**+**	"If I can't apply the logic of use that I learnt using other devices"	1
6	**+**	"If it doesn't have unconventional features"	1
7	**+**	"If its menu is like similar products"	1
8	**-**	"If it doesn't bear similarities to other products"	1
9	**+**	"If I can utilize my previous experiences"	1
10	**-**	"If it is a product with an unconventional design"	1
11	**-**	"If it has buttons and controls with unusual style of use"	1
12	**-**	"If it is a very unusual device"	1
13	**-**	"If it is a modern device"	1
14	**-**	"If its buttons contradict with their general uses"	1
15	**+**	"If I came across with a similar menu"	1
16	**-**	"If it is very different from devices that I used"	2
17	**-**	"If I am alien to the product"	1

Novelty – familiarity > familiar brand

1	**+**	"If it is a product of a brand that I am used to"	1
2	**+**	"If I used that brand's other products before"	1
3	**-**	"If it is a new brand"	1
4	**+**	"If it is not a brand preferred by everyone"	1
5	**-**	"If it is not a known, recognized brand"	1

6	**+**	“If it is the most-selling brand”	1

Novelty – familiarity > similarity with previous model

1	**+**	“If it is a new version for an existing model”	1
2	**+**	“If it resembles previous models”	1
3	**+**	“If some features are added to an old model”	1
4	**-**	“When I replaced old device with a new one”	1
5	**-**	“If I used a different model before”	1
6	**-**	“If it has many differences with a device that I used to”	1
7	**-**	“If it looks different from a model that I previously used”	1

Novelty – familiarity > diffusion

1	**+**	“If it looks familiar because it is used by many”	1
2	**-**	“If device is not commonly used”	1
3	**-**	“If it has new technologies”	1
4	**-**	“If it is a new device”	3
5	**-**	“If I am one of the first users of the product”	1
6	**-**	“If it is not a common product”	1
7	**+**	“If it is known by majority”	1

8	**-**	"If it is not widely used"	1

Affection > interest

1	**-**	"If it is not interesting"	4
2	**-**	"If it doesn't seem interesting"	2
3	**-**	"If it is a device that I was interested with"	1
4	**-**	"If it isn't in my area of interest"	8
5	**+**	"If it is in my area of interest"	4
6	**+**	"If I am quite interested in this device"	1
7	**-**	"If I lost my interest"	1
8	**-**	"If I am not interested in this product"	1

Affection > emotion

1	**+**	"If it is a product that I love"	1
2	**+**	"If it is a product that I like"	1
3	**-**	"In times when I don't like the product"	1
4	**-**	"If I was not able to get fond of the product"	1
5	**-**	"If I didn't love the product"	1

6	**-**	"If I didn't like the product"	4
7	**-**	"If I am reactive against the device"	1
8	**-**	"If I am reluctant to learn"	2
9	**+**	"If I really want to use"	1
10	**+**	"If I really want to learn"	1
11	**-**	"If I don't want to learn"	1
12	**-**	"If I don't enjoy learning "	1
13	**+**	"If I enjoy figuring it out"	1
14	**-**	"If I don't want to use"	1
15	**-**	"If I get bored of using the device"	1
16	**-**	"If I quickly got bored of using it"	1
17	**+**	"If device makes me curious"	1
18	**-**	"If I think that it is unattractive"	1
19	**+**	"If it is suitable for users"	1
20	**-**	"If it is not a product that I liked and bought"	1

Affection > visual appeal

1	**-**	"If it is not visually pleasing"	1
2	**-**	"If its color is not attractive"	1
3	**-**	"If I didn't like to look of the product""	1
4	**+**	"If it is a well-designed product"	1
5	**+**	"If it has an interesting look"	1

Usefulness > need

1	**-**	"If I think that it is not much necessary"	1
2	**-**	"If I won't use functions of the product much"	1
3	**+**	"If it has features that fit to my personal preferences"	1
4	**-**	"If I don't need the product much"	1
5	**+**	"If it will satisfy my needs"	1
6	**+**	"If it is good enough to answer my needs"	1
7	**+**	"If device is bought out of necessity"	1
8	**+**	"If it will make my daily life easier"	1
9	**+**	"If it answers my needs"	2
10	**-**	"If I had it because it is necessary"	1
11	**-**	"If I will not be able to use it in my daily life"	1
12	**+**	"If it has many functions that I will use"	1
13	**-**	"If it has many functions and features that I don't need"	1
14	**-**	"If the product is not useful for me"	2
15	**+**	"If it is useful for me"	1
16	**+**	"If device has many aspects that I am concerned with"	1
17	**+**	"If it has good functionality"	1
18	**-**	"If it doesn't have good functionality"	1
19	**+**	"If it is necessary for me to do by job better"	*
20	**+**	"If it will help me to be better in what I do"	*

* not directly expressed by respondents

Usefulness > neccesity

1	**-**	"If it is not a device that I will always use in my daily life"	1
2	**-**	"If I don't have to use that product"	1
3	**-**	"If it is not crucial for me to use it"	1
4	**+**	"If it is a device that I frequently use in my daily life"	2
5	**+**	"If I have to use the device"	1
6	**+**	"If it is a device that will affect my way of living"	1
7	**+**	"If it is a device that I frequently use"	2
8	**-**	"If I don't have to use it always"	1
9	**-**	"If I was obliged to use it" [was previously listed under urgency]	1

Usefulness > urgency

1	**+**	"If I am doomed to use that device"	1
2	**+**	"If I don't have any alternatives and should use it"	1
3	**+**	"If I am in a hurry"	1
5	**-**	"If I have to learn it fast"	1
6	**+**	"If I should urgently learn it"	1
7	**+**	"If I am in a desperate situation"	1

Ease of use [general]

1	**+**	"If it has a simple design"	1
2	**+**	"If its design is good"	1

3	**-**	"If its design is bad"	1
4	**+**	"If it is a simple device"	1
5	**+**	"If I am warned when I make a mistake"	1
6	**+**	"If textual information is provided through its screen"	1
7	**+**	"If the logic behind its menu is suitable for me"	1
8	**+**	"If it has a simple logic behind its menu"	1
9	**+**	"If there is a warning when I make a mistake"	1
10	**-**	"If its logic is not sound"	1
11	**+**	"If its menus are easy to grasp"	1
12	**+**	"If it has a clear menu"	2
13	**+**	"If its menu is easy to read"	1
14	**+**	"If it has a simple interface"	2
15	**+**	"If it is practical to use"	1
16	**+**	"If it has a simple style of use"	1
17	**+**	"If usage is clear"	1
18	**+**	"If it is an easy-to-use device"	5
19	**+**	"If it is designed simply"	1
20	**+**	"If I can reach what I want with simple steps"	1
21	**+**	"If I believe that it is simple at first sight"	1
22	**-**	"If usage is not clear"	1
23	**-**	"If it is not clear how to use it"	1
24	**-**	"If its features are not easy to use"	1

25	-	"If it has a difficult usage"	1
26	-	"If it is an impractical product"	1
27	-	"If usage of device is complex"	4
28	-	"If menu usage is hard"	1
29	-	"If it is not practical"	1
30	-	"If it is designed in a way that it is not simple"	1
31	-	"If its interface is not comprehensible"	1
32	+	"If it is easy to find the most frequently used functions"	1
33	+	"If procedure of use is easy to recall"	1
34	+	"If actions are ordered in a proper way"	1
35	-	"If I couldn't understand how it works"	1
36	-	"If I couldn't grasp the functions of its buttons"	1

Ease of use> efficiency

1	+	"If it has shortcuts"	1
2	+	"If it doesn't have shortcuts"	1
3	+	"If it has a menu that helps reaching goals"	1
4	+	"If tasks can be done with a single button"	1
5	+	"If I can quickly access what I want"	1
6	-	"If usage is full of zigzags"	1
7	-	"If one has to complete many steps during usage"	1

8	**+**	"If I can reach its features quickly"	1

Ease of use> intuitiveness

1	**-**	"If buttons have no explanations on them"	1
2	**+**	"If the functions of buttons write on them"	1
3	**-**	"If the pictures on buttons are not explicit"	1
4	**-**	"If descriptions on buttons are not similar to the ones on other devices"	1
5	**-**	"If I often have to refer to instruction manual"	1
6	**-**	"If I can't work it out with my instincts"	1
7	**+**	"If there is proper guidance while using it"	1
8	**+**	"If directions in menus are clear"	1
9	**+**	"If there are illustrative explanations in menus"	1
10	**+**	"If icons (small pictures) are used in menus"	1
11	**-**	"If it is not a device that I can work out simply by reasoning"	1
12	**+**	"If I can sort it out at first glance"	1
13	**+**	"If device has icons on it"	1
14	**+**	"If basic functions are explicitly shown on device"	1
15	**+**	"If there are signs on device"	1
16	**+**	"If I can understand how it works by looking at symbols on it"	1
17	**+**	"If device can explain itself without instruction manual"	1
18	**+**	"If icons resemble what is tried to be explained"	1

19	**-**	"If there are icons that are incomprehensible"	1
20	**-**	"If I can't understand what buttons do"	1
21	**-**	"If signs in menus are not familiar"	1
22	**-**	"If letters in menus are not familiar"	1
23	**-**	"If there are ambiguous descriptions on product"	1
24	**+**	"If usage is shown on its front face"	1
25	**+**	"If textual information on device is descriptive"	1
26	**-**	"If texts on device do not guide me"	1
27	**+**	"If information on device guide me"	1
28	**+**	"If guidance is provided during usage"	1

Ease of use> physical characteristics

1	**+**	"If it has a single button"	1
2	**+**	"If it can be used with a single button"	2
3	**-**	"If buttons have more than one function"	1
4	**-**	"If it has many buttons"	2
5	**-**	"If it has many controls"	2
6	**+**	"If it doesn't have many controls"	1
7	**+**	"If it has a wide screen"	1
8	**+**	"If functions are defined clearly with buttons"	1
9	**+**	"If there are specific buttons for specific functions"	1

10	**+**	"If it doesn't have many controls"	1
11	**-**	"If control panel is located in a hard-to-see place"	1
12	**-**	"If panel has a complex look"	1
13	**+**	"If related controls are located together"	1
14	**-**	"If I am not able to easily use it because controls are small"	1
15	**-**	"If letters and numbers are not big enough"	1

Ease of use> simplicity >structure

1	**+**	"If it doesn't have a complex menu"	3
2	**+**	"If device doesn't have a complex structure"	1
3	**+**	"If one is not required to go deep into sub menus"	1
4	**+**	"If it is not too much complicated"	2
5	**+**	"If it doesn't have many details"	1
6	**+**	"If it is not very complicated"	2
7	**-**	"If there are many variables in menus"	1
8	**-**	"If its menu is very complex"	5
9	**-**	"If device has many menus"	1
10	**-**	"If I got bored because it has many sub menus"	1
11	**-**	"If menu has many levels"	1
12	**-**	"If menus are designed so that they are very complex"	1
13	**-**	"If content in menus is excessive"	1

14	**-**	"If menus are too much complicated"	4
15	**-**	"If it is too much detailed"	1
16	**-**	"If device has complicated features"	1
17	**-**	"If device has detailed features"	1
18	**-**	"If device is complex"	4
19	**-**	"If it is a complicated device"	3
20	**-**	"If device is complex"	2
21	**+**	"If functional structure is not staged well"	1
22	**+**	"If features are not located well"	1
23	**+**	"If content is scarce"	1
24	**-**	"If it looks complex"	1

Ease of use> simplicity >number of functions

1	**+**	"If it doesn't have many functions"	2
2	**+**	"If it doesn't have too much functions"	3
3	**+**	"If it has a small number of features"	1
4	**+**	"If device has a small number of functions"	1
5	**-**	"If it has many functions"	2
6	**-**	"If it is a multi-purpose device"	1
7	**-**	"If it has many features"	4
8	**-**	"If it has many programs"	1

Ease of use> language >literal

1	**+**	“If features and functions are termed well”	1
2	**-**	“If features are badly named”	1
3	**-**	“If technical terms that are used are not easy to understand”	1
4	**-**	“If names of main menus and submenus are inconsistent”	1
5	**-**	“If menu titles are not meaningful”	1
6	**-**	“If there are incomprehensible words that are not used in daily life”	1
7	**+**	“If product speaks users’ language”	1
8	**+**	“If there is language option for menus”	1
9	**+**	“If its menus are in Turkish”	2
10	**+**	“If there are labels in Turkish”	1
11	**-**	“If I don’t know the language used in the product”	2
12	**-**	“If there is a language chaos in the product”	1
13	**+**	“If language is clear”	1
14	**-**	“If language is not neat”	1

Ease of use> language >literal

1	**+**	“If there are shapes in menus”	1

2	**+**	"If there are pictures in menus"	1
3	**+**	"If it has colorful menus"	1
4	**-**	"If there are entities in the menus that attract attention"	1
5	**-**	"If only straight black texts are used"	1

Help and support > informal help > from salespeople

1	**+**	"If I can get tips about use from where I buy the product"	1
2	**+**	"If there is someone where I buy it who teaches how to use the product"	1
3	**+**	"If explanations are provided during purchase"	2
4	**-**	"If seller doesn't help me"	1
5	**+**	"If seller shows me how to use it"	1
6	**+**	"If seller helps me"	1

Help and support > informal help > user forums

1	**+**	"If there are relevant forums about the product"	1

Help and support > informal help > to others

1	**-**	"If I have to learn the product and teach someone else"	1

2	**-**	"If I have to use the product for someone else"	1
3	**-**	"If I have to quickly install and use the product"	1

Help and support > informal help > from others

1	**+**	"If I can get info from others that use the device"	3
2	**+**	"If I have the opportunity to ask people who know the product"	1
3	**+**	"If usage is explained by someone who knows how to use it"	3
4	**+**	"If someone can briefly show how the product is used"	1
5	**+**	"When a person who know it shows me"	3
6	**+**	"If I have friend that used the product before"	1
7	**+**	"If there is someone that I can ask for help when I have problems"	1
8	**+**	"If I have the opportunity to observe someone while using the product"	1
9	**+**	"If I can get support from someone interested"	1
10	**+**	"If an acquaintance can teach me how to use it"	1
11	**+**	"If I can get help from someone that knows the product"	1
12	**-**	"If there is nobody that can support me while learning the product"	1
13	**+**	"If I can get support from people that previously used it"	1
14	**+**	"If I can get advice from people that previously used it"	1
15	**+**	"If someone who knows how to use it can show me"	1
16	**+**	"If someone who uses the product can explain"	1
17	**+**	"If an expert tells me how to use it"	1

18	**-**	"If there is nobody to help me"	3
19	**-**	"If there is nobody using it"	1
20	**-**	"If there is nobody to help me when I got stuck"	1
21	**-**	"If there is no one around to show me how to use it"	1
22	**-**	"If there is no one that I can get information"	1
23	**-**	"If there is nobody who knows the product"	1
24	**-**	"If there is nobody to guide me"	1
25	**-**	"If there is nobody to explain it in detail"	1
26	**-**	"If there is no one to tell me how to use it"	1

Help and support > formal help > instruction manual >availability

1	**+**	"If it has an instruction manual"	1
2	**+**	"With the help of instruction manual "	1
3	**-**	"If there is no instruction manual"	3
4	**-**	"If I don't have a source"	1
5	**+**	"If I can get help from its guide"	1
6	**+**	"If it has a good help menu"	1
7	**+**	"If there is a section in the instruction manual such as a "quickstart" that briefly explains how to use it"	1
8	**+**	"If there is a section in the device that show how to use it"	1
9	**+**	"If there is a CD that explains how to use it"	1

Help and support > formal help > instruction manual > characteristics

1	**+**	"If manual is plain"	1
2	**-**	"If manual has ambiguities"	1
3	**-**	"If manual is hard to comprehend"	5
4	**-**	"If information provided in the manual are not clear"	1
5	**-**	"If manual is not good"	1
6	**-**	"If manual confuses me"	1
7	**-**	"If manual is sketchy"	1
8	**-**	"If manual is not sufficient"	5
9	**-**	"If there are long explanations in the manual"	2
10	**-**	"If there are pages-long instructions in the manual"	1
11	**-**	"If manual is not clear"	4
12	**-**	"If manual is not sufficiently descriptive"	2
13	**-**	"If there are no diagrams in the manual"	2
14	**-**	"If manual is too much detailed"	1
15	**-**	"If writing style is bad"	1
16	**-**	"If some necessary information are skipped in the manual"	1
17	**-**	"If features of the product are not clearly explained"	1
18	**-**	"I don't have any source"	1
19	**-**	"If there are inconsistencies between guide and product"	1
20	**-**	"If manual and usage are inconsistent"	1
21	**-**	"If step by step instructions are not provided in the guide"	1

22	**-**	“If instructions of use are not clearly expressed in the manual”	1
23	**+**	“If I can apply exactly what it says in the manual”	1
24	**+**	“If manual briefly tells me what to do”	2
25	**+**	“If usage is described with cause-effect relations”	1
26	**-**	“If usage is not described with cause-effect relations”	1
27	**+**	“If there are simple directions in the manual”	1
28	**+**	“If there are step by step instructions in the manual”	1
29	**+**	“If manual explains what to do one by one”	1
30	**+**	“If instruction manual is practical to use”	1
31	**+**	“If manual is comprehensible enough”	1
32	**+**	“If instruction manual is illustrative”	1
33	**+**	“If I can understand the manual	2
34	**+**	“If manual is detailed enough”	2
35	**+**	“If what is described in the manual can be seen in the product”	1
36	**+**	“If manual only explains my product”	1
37	**+**	“If it has a good manual”	1
38	**+**	“If practical instructions are provided in the manual”	1
39	**+**	“If descriptions in the manual are good”	1
40	**+**	“If manual is comprehensible”	6
41	**+**	“If manual is explicit”	2
42	**+**	“If instruction manual has a plain language”	1
43	**+**	“If instruction manual is clear”	12
45	**+**	“If manual is easy to use”	1
46	**+**	“If instruction manual simply explains”	2
47	**+**	“If instruction manual very clearly explains”	3

48	**+**	“If instruction manual uses simple tables to explain”	1
49	**+**	“If manual explains with figures”	5
50	**+**	“If there are diagrams in manual”	3
51	**+**	“If there are pictures in the manual”	1
52	**-**	“If there are words in the manual that are not used in everyday language”	1
53	**-**	“If manual is written in a language that I don’t speak”	1
54	**-**	“If manual is in a foreign language”	1
55	**-**	“If instruction manual is in English” [Turkish audience]	1
56	**+**	“If there are Turkish explanations”	1
57	**-**	“If manual is not Turkish”	2
58	**+**	“If Turkish translation is successful”	1
59	**+**	“If it is translated with good Turkish”	1
60	**-**	“If manual is written in a foreign language”	4
61	**+**	“If manual is Turkish”	2
62	**+**	“If the language used is clear”	1
63	**+**	“If the language used in manual is simple”	1
64	**-**	“If technical terms are used”	1
65	**+**	“If a comprehensible written language (Turkish) is used”	1
66	**-**	“If use of language is bad”	2

Help and support > formal help > instruction manual >support services

1	**-**	“If it has no internet page”	1
2	**+**	“If it has an internet page”	1
3	**+**	“If I can get assistance from call center”	1
4	**+**	“If I can access technical service”	1
5	**+**	“If I can call customer service”	1
6	**-**	“If there is no technical service system”	1
7	**-**	“If there is no help center”	1
8	**+**	“If there is a call center”	1

Learning context and process >method

1	**+**	“If I read the manual”	5
2	**-**	“If I wasn’t able to read the manual”	2
3	**+**	“If I can do some practice”	1
4	**+**	“If I can learn with trial and error”	3
5	**-**	“If I can’t figure it out intuitively”	1
6	**-**	“When I try to learn it without reading the manual”	1
7	**-**	“If I have no chance for learning with trial and error”	1
8	**-**	“If I have to learn it theoretically”	1
9	**-**	“If I have to learn it without the actual device”	1
10	**-**	“If I have to learn it by directions, without hands-on experience”	1
11	**-**	“If I have to try everything one by one”	1
12	**-**	“If I have to read pages of instructions before using it”	2

Learning context and process >achievement

1	**-**	"If I still have problems after a couple of trials"	1
2	**-**	"If I experience problems in my first trial"	1
3	**-**	"If I am concerned of new problems, after having some problems with it"	1
4	**-**	"If I make many mistakes"	1
5	**+**	"If I feel that I am figuring it out"	1

Learning context and process >opportunities

1	**-**	"If I can only use it for short periods of time"	1
2	**-**	"If I don't have many opportunities for using it"	1
3	**+**	"If I have plenty of time for learning it"	1
4	**-**	"If I have a little time for learning it"	1
5	**+**	"If I often find the opportunity to use the product"	1
6	**+**	"If installing and disassembling the product takes too much time"	1
7	**-**	"If its charge does not last much"	1

Learning context and process >other users

1	**-**	“If there are others interfering when I try to learn it”	1
2	**-**	“If there is someone more talented next to me”	1
3	**-**	“If there is someone quicker than me”	1
4	**+**	“If I can learn faster than others around”	1
5	**-**	“If there is someone who already undertook the usage of that device”	1
6	**-**	“If I am asked to quickly install and use the device”	1
7	**+**	“If it is used before by someone else”	1
8	**+**	“If it is bought by someone else before”	1
9	**-**	“If I heard that device is complex before”	1

Breakdowns>cost

1	**-**	“If I can’t have the opportunity to try it because it is too expensive”	1
2	**-**	“If I can’t use trial and error methods because the device is too expensive”	1
3	**-**	“If risk of damaging the device is high”	1
4	**-**	“If I think that it will be damaged”	1
5	**-**	“If it breaks down easily”	1
6	**-**	“If device is prone to damage”	1
7	**-**	“If it is hard to get it fixed when it breaks down”	1
8	**-**	“If it is not possible to fix a mistake”	1
9	**-**	“If it is hard to return when I make a mistake”	1

Breakdowns>likelihood

1	**-**	"If I think that device gets easily damaged"	1
2	**-**	"If it breaks down when it is improperly used"	1
3	**-**	"If it is a delicate device"	1
4	**-**	"If I hesitate to use the product"	1
5	**-**	"If I am scared to use the product"	1
6	**-**	"If it is hard to return when a mistake is done"	1

Prior knowledge>terminology

1	**+**	"If I know what abbreviations stand for"	1
2	**+**	"If I know the terms"	1
3	**-**	"If there are many specific terms"	1
4	**-**	"If there are many abbreviations"	1

Prior knowledge>domain knowledge

1	**-**	"If I don't have the necessary knowledge"	1
2	**-**	"If I don't have the necessary background"	1
3	**-**	"If it isn't suitable for my level of knowledge"	1
4	**-**	"If I don't have prior knowledge about the product"	1

5	**-**	"If device requires extra knowledge that is beyond my experience"	1
6	**-**	"If it includes complex information"	1

APPENDIX D

Expert Review Definitions and Instructions (Sample)

Genel Etkileşim Öz Yeterliği Ölçeği – Madde Havuzu Uzman Değerlendirmesi

*Katılacağınız bu çalışma Orta Doğu Teknik Üniversitesi, Endüstri Ürünleri Tasarımı Bölümü Doktora Programı kapsamında yürütülmekte olan **tez başlığı** başlıklı araştırma için uzman görüşü almayı amaçlamaktadır.*

Araştırmada kullanılabilirlik testleri ve kullacılarla gerçekleştirilen benzeri ürün geliştirme – değerlendirme çalışmalarında kullanılmak üzere bir ölçek geliştirilmektedir. Bu tip çalışmalara yönelik bir ölçek geliştirme gereksinimi kabaca tanımlanacak olursa kullanıcılar arası 'arayüz deneyimi' çeşitliliğinin etkili olduğu niceliksel ve niteliksel 'performans' farklarından kaynaklanmaktadır.

> *Ölçeğin*(1) amacı *çalışmalara*(2) katılan kullanıcıların *dijital arayüzlere*(3) sahip ürünleri konusundaki *öz yeterliklerini* (4) ölçmektir.

Tanımlar
Uzman görüşünün çalışmada tanımlanan işlevini yerine getirmesi önündeki en büyük engel tanımlanan kavramların belirgin bir şekilde uzmanlara aktarılamaması olarak görüldüğünden yukarıda kullanılan kavramlar açık bir şekilde tarif edilmeye çalışılmıştır.

1 - 'Ölçek'
Bu çalışmada ölçek, bir çok sorunun (madde) tek bir kavramı farklı açılardan irdelemek amacıyla katılımcıya yöneltilmesi ve katılımcının bir konuyu tutum, yaklaşım, yakınlık gibi psikolojik kavramlar açısından değerlendirmesi olarak tanımlanmıştır. Dolayısıyla, ölçek maddeleri doğru veya yanlış cevapları olmayan ve katılımcının kendisiyle ilgili yargılarını anlamaya çalışmak amacıyla geliştirilmiş ifadelerden (madde gövdesi) ve cevaplama formatından oluşmaktadır. Çalışma sonucunda geliştirilecek olan ölçeğin kısa bir açıklamanın ardından katılımcılar tarafından bir anketör veya gözlemci olmadan cevaplanması düşünülmektedir.

2 - 'Çalışmalar'
Kapsama giren çalışmalarda genel anlamda katılımcının ilk kez karşılaştığı bir arayüzü daha önceden belirlenmiş kullanım senaryoları doğrultusunda deneyimlemesi ve bu sırada verimlilik – etkinlik değerlerinin ölçülmesi, yaşanan problemlerin gözlenmesi ve ilişkili konulardaki fikirlerinin alınmasına dayalı bir yöntem izlenmektedir. Temel tanım bu olmakla beraber katılımcıların yeni ürünlerle karşılaştığı ya da ürün fikirleri geliştirmeleri, değerlendirmeleri istendiği durumlar da 'çalışma' tanımına girerek ölçeğin kullanım alanını genişletmektedir.

3 - 'Dijital arayüz'
'Dijital' tanımlaması arayüz teknolojisine işaret etmekten çok arayüz tipine ve karmaşa seviyesine referans vermektedir. Tanımlanmaya çalışılan arayüz göstergeler ve tuşlar aracılığıyla kullanılan derin arayüzlü ürünleri kapsamaktadır. Ancak katılımcılara verilecek açıklamalarda 'elektronik alet' anahtar sözcüğü kullanılmış daha sonra 'dijital ekranları veya elektronik panelleri bulunan, menüler aracılığıyla kullanılan dijital ürünler' olarak tanımlanmıştır.

4 - 'Öz yeterlik'
Katılımcılara açıklanmamakla beraber ölçek öz yeterlik kavramı üzerinde yoğunlaşmaktadır. Öz yeterlik kavramı A. Bandura tarafından ortaya atılmış olup bireylerin belirli bir işi yapma konusunda kendilerine duydukları güven olarak tanımlanabilir. Bandura'ya göre öz yeterlik belirli bir işi tamamlayabilmek için gereken becerilere sahip olmanın ötesinde bir kavram olarak tanımlanabilmektedir[1]. Buna göre öz yeterlik bir işin nasıl sonuçlanacağına ilişkin bir beklenti değil bireyin herhangi bir işte belirli bir seviyede performans gösterebilmesine ilişkin öz inancıdır. Ancak öz yeterlik sadece bireyin yapabildiklerinin bir yansıması değil işlevsel bir mekanizmadır. Bandura'ya göre öz yeterlik inancı bireyin gelecekteki performansına ilişkin tahminler yürütmeye yaradığı gibi bireyin hangi tür seçimler yapacağını da belirleyerek bir ölçüde kişisel gelişime yön verir.

Genel Etkileşim Öz Yeterliği Kavramı

Genel etkileşim öz yeterliği kavramı bireylerin yeni elektronik aletleri öğrenme konusundaki öz yeterlik algıları olarak ele alınmaktadır. Her ne kadar bu tanım içersinde öğrenme uzun dönemli bir süreci kapsamamaktaysa da ölçeğin kapsadığı alanı çok daraltmamak için öğrenme genel anlamıyla kabul edilmiştir.

> Genel Etkileşim Öz Yeterliği yeni bir elektronik aletle etkileşime girme, etkileşim sırasında karşılaşılabilecek yeni durumlara adapte olabilme becerisi olarak tanımlanmaktadır.

Değerlendirme Kriterleri

Yapılan tanımlamalar ışığında maddelerin değerlendirilmesine geçilmeden önce ölçek maddelerinin hangi kriterlere göre değerlendirileceği bu bölümde kısaca ele alınmıştır. Bu kriterler sadece literatürden derlenmiş olan prensipler hakkında fikir vermek için sunulmuştur. Uzmanlar maddeleri uygun gördükleri şekilde, farklı kriterlere göre değerlendirebilirler. Ancak yapılacak olan değerlendirmelerin aşağıda verilen grup tanımlamalarına uygun olması gerekmektedir. Uzmanlar bu gruplar dışında kalan, maddeleri farklı açılardan değerlendiren yorumlarını da ayrıca not alabilirler.

Değerlendirme kriterleri anlatım ve içerik olmak üzere iki ana grup altında toplanmıştır.

Anlatım - Maddenin yazım - anlatım – iletişimsel nitelikler açısından değerlendirilmesi
İçerik ne olursa olsun madde gövdelerinin katılımcılar tarafından rahatlıkla anlaşılabilecek nitelikte olması gerekmektedir.

1) Maddelerde kullanılan sözcüklerde teknik terimler ve sadece belirli bir kesimin anlayabileceği ya da farklı anlamlara gelebilecek jargon kullanılmamalıdır.

2) Anlatımı düzgün olsa da dikkat dağılmasına neden olabilecek uzun maddelerden kaçınılmalıdır.

3) Maddeler katılımcıları olumlu ya da olumsuz yönde koşullandırabilecek içermemelidir.
   Örn. "Her çağdaş insan gibi kitap okumayı severim."

[1] Bandura, A. (1986). *Social foundations of thought and action*. Londra: Prentice Hall.

4) Birden fazla yargı içeren, tek bir cevapla yanıtlanamayacak sorulardan kaçınmak gerekmektedir.
   Örn. "Kitap okumayı ve müzik dinlemeyi severim"
5) Çift olumsuz içeren cümleler kullanılmamalıdır.
   Örn. "Doğru olan kitap okumayı sevmemek değildir"

6) İçeriği ne olursa olsun zayıf veya çok kuvvetli argümanlar, ya da herkesin kabul edeceği ya da reddedeceği argümanlardan kaçınılmalıdır.

Not. Birbirine çok benzer şekilde yazılmış maddelerin elenmesinin veriye dayalı olarak yapılması yöntemsel açıdan daha uygun olacağından benzer şekilde yazılmış maddeler ilk aşamada korunmalıdır.

İçerik – Maddenin amaca uygunluğu açısından değerlendirilmesi

Maddelerin hedeflenen kavramla ilgili içeriğe sahip olması gerekmektedir.

1) Maddeler cevap vermeyi zorlaştıracak şekilde birçok olayı - durumu kapsayacak veya genel kavramlar içerecek şekilde yazılmamalıdır.

2) Maddeler çok spesifik bir olayı – durumu ele almamalıdır.

3) Madde içerikleri hedeflenen kavramı temsil eder nitelikte olmalıdır.

4) Maddeler katılımcı profili konusunda ön yargılı olmamalıdır.
   Örn. "Okuduğum kitaplar İngilizce terimler içeriyorsa anlayamıyorum" – Katılımcının cevabı İngilizce bilip bilmediğine göre değişecektir.

5) Maddeler katılımcıların deneyimlememiş olabileceği durumlar üzerine kurulmamalıdır.
   Örn. "Bilimsel makaleleri okumak gazete okumak kadar kolay" – Katılımcıların bir kısmı bilimsel makale okumamış olabileceğinden bu maddeyi cevaplayamayacaktır.

6) Madde içeriklerinde bir çok konuda (cinsel, dini, ırka dayalı, ideolojik v.b.) ayrımcılıktan kaçınılmalıdır.

Not. Tam olarak Genel Etkileşim Öz Yeterliği kavramı içine girmeyen ancak yakın anlamlar içeren maddeler veriye dayalı olarak elenmek üzere bu aşamada korunabilir.

Değerlendirme

Verilen bu tanımlamalar ışığında lütfen değerlendirme formlarını doldurunuz. Değerlendirmeler esas olarak 2 farklı puanlamayı gerektirmektedir. Maddeleri önce yukarıda tanımlandığı şekliyle anlatım açısından, sonra içerik açısından 1-9 arasında puan vererek değerlendiriniz. Daha sonra, ayrılan bölüme değerlendirmede etkili olan düşüncelerinizi belirtebilirsiniz. Yapmak istediğiniz düzeltmeleri de bu bölümde yapabilirsiniz. Formun sonunda yer alan kısımda ölçeğe eklenmesi gerektiğini düşündüğünüz maddeleri yazabilirsiniz. Genel öneriler kısmında ölçekle ilgili her konudaki genel yorumlarınızı, önerilerinizi yazabilirsiniz.

Teşekkürler...

APPENDIX E

GISE-S EXPERT REVIEW FORM (SAMPLE PAGES)

Genel Etkileşim Öz Yeterliği Ölçeği – Madde Havuzu Uzman Değerlendirmesi

Katılacağınız bu çalışma Orta Doğu Teknik Üniversitesi, Endüstri Ürünleri Tasarımı Bölümü Doktora Programı kapsamında yürütülmekte olan bir araştırma için uzman görüşü almayı amaçlamaktadır.

Araştırma kullanılabilirlik testleri ve kullanıcılarla gerçekleştirilen benzeri ürün geliştirme – değerlendirme çalışmalarında kullanılmak üzere bir ölçek geliştirilmesini amaçlamaktadır. Bu tip çalışmalara yönelik bir ölçek geliştirme gereksinimi genel anlamda kullanıcılar arası 'arayüz deneyimi' çeşitliliğinin etkili olduğu niceliksel ve niteliksel 'performans' farklarından kaynaklanmaktadır.

Ölçeğin(1) amacı *çalışmalara*(2) katılan kullanıcıların *dijital arayüzlere*(3) sahip ürünleri konusundaki *öz yeterliklerini* (4) ölçmektir.

Tanımlar
Uzman görüşünün çalışmada tanımlanan işlevini yerine getirmesi önündeki en büyük engel tanımlanan kavramların belirgin bir şekilde uzmanlara aktarılamaması olarak görüldüğünden yukarıda kullanılan kavramlar açık bir şekilde tarif edilmeye çalışılmıştır.

1 - 'Ölçek'
Bu çalışmada ölçek, bir çok sorunun (madde) tek bir kavramı farklı açılardan irdelemek amacıyla katılımcıya yöneltilmesi ve katılımcının bir konuyu tutum, yaklaşım, yakınlık gibi psikolojik kavramlar açısından değerlendirmesi olarak tanımlanmıştır. Dolayısıyla, ölçek maddeleri doğru veya yanlış cevapları olmayan ve katılımcının kendisiyle ilgili yargılarını anlamaya çalışmak amacıyla geliştirilmiş ifadelerden (madde gövdesi) ve cevaplama formatından oluşmaktadır. Çalışma sonucunda geliştirilecek olan ölçeğin kısa bir açıklamanın ardından katılımcılar tarafından bir anketör veya gözlemci olmadan cevaplanması düşünülmektedir.

2 - 'Çalışmalar'
Kapsama giren çalışmalarda genel anlamda katılımcının ilk kez karşılaştığı bir arayüzü daha önceden belirlenmiş kullanım senaryoları doğrultusunda deneyimlemesi ve bu sırada verimlilik – etkinlik değerlerinin ölçülmesi, yaşanan problemlerin gözlenmesi ve ilişkili konulardaki fikirlerinin alınmasına dayalı bir yöntem izlenmektedir. Temel tanım bu olmakla beraber katılımcıların yeni ürünlerle karşılaştığı ya da ürün fikirleri geliştirmeleri, değerlendirmeleri istendiği durumlar da 'çalışma' tanımına girerek ölçeğin kullanım alanını genişletmektedir.

3 - 'Dijital arayüz'
'Dijital' tanımlaması arayüz teknolojisine işaret etmekten çok arayüz tipine ve karmaşa seviyesine referans vermektedir. Tanımlanmaya çalışılan arayüz göstergeler ve tuşlar aracılığıyla kullanılan derin arayüzlü ürünleri kapsamaktadır. Ancak katılımcılara verilecek açıklamalarda 'elektronik alet' anahtar sözcüğü kullanılmış daha sonra 'dijital ekranları veya elektronik panelleri bulunan, menüler aracılığıyla kullanılan dijital ürünler' olarak tanımlanmıştır.

4) Birden fazla yargı içeren, tek bir cevapla yanıtlanamayacak sorulardan kaçınmak gerekmektedir.
   Örn. "Kitap okumayı ve müzik dinlemeyi severim"
5) Çift olumsuz içeren cümleler kullanılmamalıdır.
   Örn. "Doğru olan kitap okumayı sevmemek değildir"

6) İçeriği ne olursa olsun zayıf veya çok kuvvetli argümanlar, ya da herkesin kabul edeceği ya da reddedeceği argümanlardan kaçınılmalıdır.

Not. Birbirine çok benzer şekilde yazılmış maddelerin elenmesinin veriye dayalı olarak yapılması yöntemsel açıdan daha uygun olacağından benzer şekilde yazılmış maddeler ilk aşamada korunmalıdır.

İçerik – Maddenin amaca uygunluğu açısından değerlendirilmesi

Maddelerin hedeflenen kavramla ilgili içeriğe sahip olması gerekmektedir.

1) Maddeler cevap vermeyi zorlaştıracak şekilde birçok olayı - durumu kapsayacak veya genel kavramlar içerecek şekilde yazılmamalıdır.

2) Maddeler çok spesifik bir olayı – durumu ele almamalıdır.

3) Madde içerikleri hedeflenen kavramı temsil eder nitelikte olmalıdır.

4) Maddeler katılımcı profili konusunda ön yargılı olmamalıdır.
   Örn. "Okuduğum kitaplar İngilizce terimler içeriyorsa anlayamıyorum" – Katılımcının cevabı İngilizce bilip bilmediğine göre değişecektir.

5) Maddeler katılımcıların deneyimlememiş olabileceği durumlar üzerine kurulmamalıdır.
   Örn. "Bilimsel makaleleri okumak gazete okumak kadar kolay" – Katılımcıların bir kısmı bilimsel makale okumamış olabileceğinden bu maddeyi cevaplayamayacaktır.

6) Madde içeriklerinde bir çok konuda (cinsel, dini, ırka dayalı, ideolojik v.b.) ayrımcılıktan kaçınılmalıdır.

Not. Tam olarak Genel Etkileşim Öz Yeterliği kavramı içine girmeyen ancak yakın anlamlar içeren maddeler veriye dayalı olarak elenmek üzere bu aşamada korunabilir.

Değerlendirme

Verilen bu tanımlamalar ışığında lütfen değerlendirme formlarını doldurunuz. Değerlendirmeler esas olarak 2 farklı puanlamayı gerektirmektedir. Maddeleri önce yukarıda tanımlandığı şekliyle anlatım açısından, sonra içerik açısından 1-9 arasında puan vererek değerlendiriniz. Daha sonra, ayrılan bölüme değerlendirmede etkili olan düşüncelerinizi belirtebilirsiniz. Yapmak istediğiniz düzeltmeleri de bu bölümde yapabilirsiniz. Formun sonunda yer alan kısımda ölçeğe eklenmesi gerektiğini düşündüğünüz maddeleri yazabilirsiniz. Genel öneriler kısmında ölçekle ilgili her konudaki genel yorumlarınızı, önerilerinizi yazabilirsiniz.

Teşekkürler...

Değerlendirmeyi yapan ________________________________

No	Madde	Anlatım (1-9)	İçerik (1-9)	No
1	Daha önce kullandığım tür bir alet değilse			1
2	Daha önceden kullanmadığım bir tür aletse			2
3	Daha önce aynı işe yarayan bir aleti kullanmadıysam			3
4	Daha önce karşılaşmadığım bir aletse			4
5	Daha önceden kullandığım aletlere benzemiyorsa			5
6	Kullanımı önceden bildiğim aletlere benzemiyorsa			6
7	Sık sık kullandığım aletlere benzemiyorsa			7
8	Diğer aletlerden bildiğim kullanım şeklini uygulayamıyorsam			8
9	Çok değişik özelliklere sahipse			9
10	Menüsü aynı tür aletlerin menüsüne benzemiyorsa			10
11	Diğer aletlere benzemiyorsa			11
12	Önceki aletlerden kazandığım tecrübeyi kullanamıyorsam			12
13	Daha önce benzer bir menüyle karşılaşmışsam			13
14	Daha önce kullandığım aletlerden çok farklıysa			14
15	Bana yabancı bir aletse			15
16	Alıştığım bir markaya ait değilse			16
17	Aynı markaya ait başka alet kullanmamışsam			17
18	Herkes tarafından tercih edilen bir markaya ait değilse			18
19	Alıştığım bir aletin yeni modeli değilse			19

Note. The rest of the items were provided in Appendix E

APPENDIX F

ITEMS IN THE FIRST ITEM POOL – ENGLISH AND TURKISH (EXPERT REVIEW PHASE)

WARNING: The expressions listed below were not translated using a systematic procedure and no data was collected in order to provide an English version of GISE-S. Therefore, following item stems should not be used for item generation or data collection.

No	Item
1	Daha önce kullandığım tür bir alet değilse If it is not a type of device that I used before
2	Daha önceden kullanmadığım bir tür aletse If it is a type of device that I didn’t use before
3	Daha önce aynı işe yarayan bir aleti kullanmadıysam If it is not a type of device that I uses before
4	Daha önce karşılaşmadığım bir aletse If it is a type of device that I didn’t use before
5	Daha önceden kullandığım aletlere benzemiyorsa If it doesn’t resemble devices that I used before
6	Kullanımı önceden bildiğim aletlere benzemiyorsa If its use isn’t similar to devices that I used before
7	Sık sık kullandığım aletlere benzemiyorsa If it is not similar to a device that I often use
8	Diğer aletlerden bildiğim kullanım şeklini uygulayamıyorsam If I can’t apply the style of use that I learnt using other devices

9 Çok değişik özelliklere sahipse
If it has unconventional features

10 Menüsü aynı tür aletlerin menüsüne benzemiyorsa
If its menu is not like similar products

11 Diğer aletlere benzemiyorsa
If it doesn’t bear similarities to other products

12 Önceki aletlerden kazandığım tecrübeyi kullanamıyorsam
If I can’t utilize my previous experiences

13 Daha önce benzer bir menüyle karşılaşmışsam
If I didn’t come across with a similar menu

14 Daha önce kullandığım aletlerden çok farklıysa
If it is very different from devices that I used

15 Bana yabancı bir aletse
If I am alien to the product

16 Alıştığım bir markaya ait değilse
If it is a product of a brand that I am used to

17 Aynı markaya ait başka alet kullanmamışsam
If I used that brand’s other products before

18 Herkes tarafından tercih edilen bir markaya ait değilse
If it is not a brand preferred by everyone

19 Alıştığım bir aletin yeni modeli değilse
If it is not a new version for an existing model I got used to

20 Daha önceki modelleriyle benzerlik taşımıyorsa
If it does not resemble previous models

21 Daha önce alıştığım aletle arasında çok fark varsa
If it has many differences with a device that I used to

22 Aletin kullanımı yaygın değilse
If device is not commonly used

23 Yeni teknolojiler içeriyorsa
If it has new technologies

24 Çok yeni bir aletse
If it is a new device

25 Aletin ilk kullanıcılarındansam
If I am one of the first users of the product

26 Yaygın olmayan bir aletse
If it is not a common product

27 Kullanımı yaygın olmayan bir aletse
If it is not widely used

28 Alet ilgimi çekmemişse
If it is not interesting

29 Alet bana İlgi çekici gelmediyse
If it doesn't seem interesting

30 Çok ilgilenmediğim bir aletse
If it is a device that I was not interested with

31 Alet ilgi alanıma girmiyorsa
If it isn't in my area of interest

32 Alete karşı ilgim fazla değilse
If I am not much interested in this device

33 Sevdiğim tür bir alet değilse
If it is not a product that I love

34 Hoşlandığım bir alet değilse
If it is not a product that I like

35 Alete fazla ısınamadıysam
If I was not able to get fond of the product

36 Aleti fazla sevmediysem
If I didn’t love the product

37 Aletten çok hoşlanmamışsam
If I didn’t like the product

38 Kullanmayı gerçekten istemiyorsam
If I do not really want to use

39 Öğrenmeyi gerçekten istemiyorsam
If I don’t want to learn

40 Öğrenmekten zevk almıyorsam
If I don’t enjoy learning

41 Nasıl kullanıldığını çözmek hoşuma gitmiyorsa
If I don’t enjoy figuring it out

42 Aleti kullanmak beni sıkıyorsa
If I get bored of using the device

43 Öğrenmekten çabuk sıkıldığım bir aletse
If I quickly get bored of using it

44 Alet bende merak uyandırmıyorsa
If device does not make me curious

45 Alet bana itici geliyorsa
If I think that it is unattractive

46 Severek aldığım bir alet değilse
If it is not a product that I liked and bought

47 Çok gerek görmediğim bir aletse
If I think that it is not much necessary”

48 Özelliklerini çok fazla kullanmayacaksam
If I won’t use functions of the product much

49 Fazla ihtiyaç duymadığım bir aletse
If I don’t need the product much

50 İhtiyaçlarımı karşılayacak bir alet değilse
If it will not satisfy my needs

51 İhtiyaçlarıma cevap verecek nitelikte değilse
If it is not good enough to answer my needs

52 Alet ihtiyaçtan alınmamışsa
If device is not bought out of necessity

53 Günlük hayatımı kolaylaştıracak bir alet değilse
If it will not make my daily life easier

54 İhtiyaçlarıma cevap vermiyorsa
If it does not answer my needs

55 İhtiyaçtan ötürü alınmış bir alet değilse

If it is not a device that is bought out of necessity

56 Günlük hayatta kullanabileceğim bir alet değilse
If I will not be able to use it in my daily life

57 Kullanmayacağım özellikleri varsa
If it has many functions that I won’t use

58 İşime yaramayacak özellikleri çoksa
If it has many features that I do not need

59 İşime yaramayacak bir aletse
If the product is not useful for me

60 İşimi daha iyi yapmam için gerekli bir alet değilse
If it is not necessary for me to do by job better

61 Yaptığım işleri daha iyi yapmamı sağlayacaksa

If it will not help me to be better in what I do

62 Özelliklerinin çoğu işime yaramıyorsa
If I will not need many of its features

63 Günlük hayatta sürekli kullanacağım bir alet değilse
If it is not a device that I will always use in my daily life

64 Kullanmak zorunda olduğum bir alet değilse

If it is not a device that I have to use

65 Aleti kullanmam gerekli değilse

If I don’t have to use that device

66 Sıkça kullanıdığım bir alet değilse
If it is not a device that I frequently use

67 Sürekli kullanmam gerekmiyorsa
If I don’t have to use it always

68 Aleti kullanmaya mecbur değilsem
If I was obliged to use it

69 Aleti kullanmam şart değilse
If I am not doomed to use that device

70 Basit bir alet değilse

If it is not a simple device

71 Menüsü bana ters geliyorsa
If the logic behind its menu is not suitable for me

72 Menü kullanımı kolay değilse

If menu usage is not easy

73 Menüsü açık - net değilse
If it does not have a clear menu

74 Basit bir kullanımı yoksa
If it does not have a simple style of use

75 Nasıl kullanılacağı açık değilse
If usage is not clear

76 Kolay kullanılabilen bir alet değilse
If it is not an easy-to-use device

77 Basit adımlarla istediğime ullaşmam mümkün değilse
If I can not reach what I want with simple steps

78 İlk görüşte bana zor göründüyse
If I believe that it is hard at first sight

79 Kullanım açık değilse
If usage is not clear

80 Nasıl kullanılacağı net değilse

If it is not clear how to use it

81 Kullanımı zor bir aletse
If it has a difficult usage

82 Aletin kullanımı karışıksa
If usage of device is complex

83 Çok kullanılan özellikleri kolay bulunamıyorsa
If it is not easy to find the most frequently used functions

84 Kullanım aşamaları akılda kalıcı değilse
If procedure of use is not easy to recall

85 Çalışma biçimini kavrayamadıysam
If I couldn’t understand how it works

86 Tuşların ne işe yaradığı açık değilse
If I couldn’t grasp the functions of its buttons

87 Hızlı bir şekilde istediğime ulaşamıyorsam
If I cannot quickly access what I want

88 Kullanımı dolambaçlı olursa
If usage is full of zigzags

89 Kullanım sırasında bir sürü aşamadan geçmek gerekiyorsa
If one has to complete many steps during usage

90 Özelliklere hemen ulaşamıyorsam
If one has to complete many steps during usage

91 Tuşların açıklamaları yoksa
If buttons have no explanations on them

92 Tuşların üstünde ne işe yaradıkları yazılı değilse
If the functions of buttons doew not write on them

93 Tuşların üstündeki resimler belirgin değilse
If pictures on buttons are not explicit

94 Tuşların üstündeki açıklamalar diğer aletlerden farklıysa
If descriptions on buttons are not similar to the ones on other devices

95 Sık sık kılavuza başvurmam gerekiyorsa
If I often have to refer to instruction manual

96 İç güdülerime dayanarak çözmem mümkün değilse
If I can't work it out with my instincts

97 Kullanım sırasında yönlendirmeler yoksa
If there is no proper guidance while using it

98 Menülerde açıklamalar net değilse
If directions in menus are not clear

99 Menülerde açıklayıcı bilgiler yoksa
If there are no illustrative explanations in menus

100 Mantık yürüterek çözebileceğim bir alet değilse
If it is not a device that I can work out simply by reasoning

101 İlk bakışta nasıl kullanılacağını anlayamadıysam
If I cannot understand how it works by looking at symbols on it

102 Temel özelliklerin nasıl kullanılacağı açık değilse
If basic functions are not easy to use

103 Kılavuza ihtiyaç duymadan alet kendi kendini anlatamıyorsa
If device can not explain itself without instruction manual

104 Anlaşılmayan resimler-semboller varsa
If there are icons that are incomprehensible

105 Tuşların ne işe yaradığı anlaşılmıyorsa

If I cannot understand what buttons do

106 Aletin üstünde belirsiz açıklamalar olursa
If there are ambiguous descriptions on product

107 Kullanım şekli aletin üstünde gösterilmiyorsa
If usage is not shown on its front face

108 Aletin üzerindeki yazılar yönlendirici değilse
If textual information on device is not descriptive

109 Aletin üstünde yönlendirici bilgiler yoksa
If information on device does not guide me

110 Kullanım sırasında yönlendirici bilgiler verilmiyorsa
If guidance is not provided during usage

111 Tuşlar birden fazla işe yarıyorsa
If buttons have more than one function

112 Çok fazla tuşu varsa
If it has many buttons

113 Menüsü çok karışıksa
If it has a complex menu

114 Alet karmaşık bir yapıya sahipse
If device has a complex structure

115 Menülerde çok fazla değişken varsa
If there are many variables in menus

116 Menüsü çok karışıksa
If its menu is very complex

117 Alette çok menü varsa
If device has many menus

118 Fazla alt menüsü varsa
If it has many sub menus

119 Menüler çok karışık yapılmışsa
If menus are designed so that they are very complex

120 Menülerin içeriği çoksa
If content in menus is excessive

121 Menüler çok karmaşıksa
If menus are too much complicated

122 Alet çok karmaşık özelliklere sahipse
If device has complicated features

123 Alet karmaşıksa
If device is complex

124 Çok fazla özelliğe sahipse
If device has many features

125 Çok özelliği varsa
If it has many features

126 Çok amaçlı bir aletse
If it is a multi-purpose device

127 Özellikler iyi adlandırılmamışsa
If features are not properly named

128 Kullanılan teknik kelimeler anlaşılmaz olursa
If technical terms that are used are not easy to understand

129 Üstünde anlaşılmayan sözcükler varsa
If there are incomprehensible words on it

130 Tuşların üstünde bilmediğim dilde yazılar varsa
If there there are labels on buttons in a language that I do not speak

131 Alette bilmediğim bir dil kullanılıyorsa
If I don’t know the language used in the product

132 Alette kullanılan dil açık değilse
If language is clear

133 Satın aldığım yerde öğreten biri yoksa
If there is nobody where I buy it that teaches how to use the product

134 Satılırken açıklayıcı bilgi verilmezse
If explanations are not provided during purchase

135 Satan yer yardımcı olmazsa
If seller does not help me (?)

136 Satıcı nasıl kullanacağımı göstermezse
If seller does not show me how to use it

137 Satış elemanı yardımcı olmazsa

If seller does not help me

138 Aleti kullananlardan bilgi alamıyorsam
If I cannot get info from others that use the device

139 Bilen kişilere sorma şansım yoksa
If I do not have the opportunity to ask people who know the product

140 Bilen biri tarafından kullanım anlatılmazsa
If usage is not explained by someone who knows how to use it

141 Nasıl kullanıldığını özetleyebilecek biri yoksa
If there is no one that can briefly show how the product is used

142 Kullanımı gösterecek biri yoksa
If there is no one to show how to it

143 Aleti daha önce kullanmış bir arkadaşım yoksa
If I do not have a friend that used the product before

144 Zorlandığımda yardım alabileceğim biri yoksa
If there is no one that I can ask for help when I have problems

145 Kullanabilen birini gözlemleme şansım yoksa
"If I do not have the opportunity to observe someone while using the product"

146 Aleti bana öğretecek bir tanıdık yoksa
If there is no acquaintance who can teach me how to use it

147 Bilen birinden yardım alamıyorsam
If I cannot get help from someone that knows the product

148 Öğrenmemi destekleyecek biri yoksa
If there is nobody that can support me while learning the product

149 Daha önce kullananlardan destek alamıyorsam
If I cannot get support from people that previously used it

150 Daha önce kullananlara danışma fırsatım yoksa
If I cannot get advice from people that previously used it

151 Kullanımı bilen bir uygulamalı olarak anlatmazsa
If someone who knows how to use it does not show me

152 Yardım alabileceğim kimse yoksa
If there is nobody to help me

153 Çevremde kullanan başka insanlar yoksa
If there is nobody using it

154 Takıldığım zaman yardım edecek kimse yoksa
If there is nobody to help me when I got stuck

155 Kullanımı gösterecek kişiler yoksa
If there is no one around to show me how to use it

156 Bilgi alabileceğim kimse yoksa
If there is no one that I can get information

157 Çevremde aleti bilen biri yoksa

If there is nobody who knows the product

158 Yönlendirecek biri yoksa
If there is nobody to guide me

159 Detaylı şekilde anlatacak biri yoksa
If there is nobody to explain it in detail

160 Kılavuzu yoksa
If it does not have an instruction manual

161 İyi bir yardım menüsüne sahip değilse
If it does not have a good help menu

162 Kılavuzda kullanımı kısaca anlatan bir bölüm yoksa
If there is not a section in the instruction manual such as a “quickstart” that briefly explains how to use it

163 Alet içinde kullanımı öğreten bir bölüm yoksa
If there is not a section in the device that show how to use it

164 Kılavuz anlaşılamıyorsa
If manual is hard to comprehend

165 Kılavuzda verilen bilgiler net değilse
If information provided in the manual are not clear

166 Kılavuz iyi değilse
If manual is not good

167 Kılavuz yetersizse
If manual is not sufficient

168 Kullanım kılavuzunda uzun anlatımlar varsa
If there are long explanations in the manual

169 Kılavuzda sayfalar dolusu açıklamalar varsa
If there are pages-long instructions in the manual

170 Kılavuz açık değilse
If manual is not clear

171 Kılavuz yeterince açıklayıcı değilse
If manual is not sufficiently descriptive

172 Kılavuzda gerekli bilgiler yoksa
If some necessary information are skipped in the manual

173 Kılavuzda kullanım adım adım anlatılıyorsa
If step by step instructions are not provided in the guide

174 Kullanım kılavuzu yeterince anlaşılır değilse
If manual is not comprehensible enough

175 Kullanım kılavuzu açıklayıcı değilse
If instruction manual is not illustrative

176 Kullanım kılavuzunda yalın bir dil yoksa
If instruction manual does not have a plain language

177 Kullanım kılavuzu açık değilse
If instruction manual is not clear

178 Kullanım kılavuzunda günlük dilde kullanılmayan sözcükler bulunuyorsa

If there are words in the manual that are not used in everyday language

179 Kılavuz bilmediğim bir dilde yazılmışsa
If manual is written in a language that I don't speak

180 Kılavuzda teknik terimler kullanılıyorsa
If technical terms are used

181 Teknik servisten telefonla yardım almak mümkün değilse
If I cannot get assistance from call center

182 Kılavuzu hiç okuma şansı bulamadıysam
If I wasn't able to read the manual

183 İstediğim kadar deneme yapma şansım yoksa
If I don't have many opportunities for using it

184 Herşeyi tek tek denemek zorunda kalıyorsam
If I have to try everything one by one

185 Kullanabilmek önce sayfalarca kılavuz okumam gerekiyorsa

If I have to read pages of instructions before using it

186 Bir kaç kez kullandığımda hala sorun yaşıyorsam
If I still have problems after a couple of trials

187 İlk kullanımda sorun yaşarsam
If I experience problems in my first trial

188 Kullanırken çok hata yapıyorsam
If I make many mistakes

189 Çözmeye başladığımı hissedemiyorsam
If I do not feel that I am figuring it out

190 Alete az zaman ayırabiliyorsam
If I can only use it for short periods of time

191 Aleti sıkça kullanma fırsatı bulamıyorsam
If I don't have many opportunities for using it

192 Öğrenmeye çalışırken yanımda bana müdahale eden biri olursa

If there are others interfering when I try to learn it

193 Başkaları yanımdayken önce ben çözemiyorsam
If I am the first to figure it out while others are around

194 Yanımda zaten o aleti kullanmayı üstlenmiş biri varsa

If there is someone who already undertook the usage of that device

195 Aletin karışık olduğunu daha önce birinden duyduysam

If I heard that device is complex before

196 Denerken aletin bozulma ihtimali varsa
If risk of damaging the device is present

197 Yanlış yaptığımda geri dönüş yoksa
If it is hard to return when I make a mistake

198 Hata yapıldığında başa dönmek zorsa
If it is hard to return when I make a mistake

199 Çabuk arızalanacak bir alet olduğunu düşünüyorsam
If I think that device gets easily damaged

200 Kullanmaya çekindiğim bir aletse

If I hesitate to use the product

201 Yanlış kullanıldığında başa dönmek zorsa

If it is hard to return when a mistake is done

202 Alette kullanılan kısaltmaların ne anlama geldiğini bilmiyorsam

If I do not know what abbreviations stand for

203 Kullanılan terimlerin ne anlama geldiğini bilmiyorsam

If I do not know the terms

204 Çok fazla özel terim kullanılıyorsa

If there are many specific terms

205 Çok fazla kısaltma kullanılıyorsa

If there are many abbreviations

206 Gerekli bilgiye sahip değilsem
If I don’t have the necessary knowledge

207 Daha önceden alet hakkında bilgim yoksa

If I don’t have the necessary background

208 Alet bilgi birikimim dışında bilgi gerektiriyorsa
If it isn’t suitable for my level of knowledge

209 Çok karışık bilgi içeriyorsa
If it includes complex information

210 İyi düşünülerek yapılmamış bir alet değilse
If it is not a well-thought device

211 Menüsü kötü yapılmışsa

If its menu is badly designed

212 Menüleri kolay kullanıma göre yapılmadıysa
If its menus are not designed for ease of use

213 Kullanım kolaylığı düşünülmeden yapılmış bir aletse

If the device is done without considering ease of use

214 Bilmediğim bir konuyla ilgliyse

If it is about something I do not know

215 Zor kontrol edilen bir aletse

If it is a device that is hard to control

216 Aletle yapılabilecek çok şey varsa

If there is much to do with the device

217 Kullanmadan önce bir sürü ayar yapmak gerekiyorsa

If there is much to do before using it

218 İlk kez açıldığında ayarlanması gereken çok şey varsa

If there is much to adjust when it is operated for the first time

219 Yaptıklarımın doğru mu yanlış mı olduğunu anlamakta zorlanıyorsam

If I can hardly understand whether the things I did are right or wrong

220 Hangi işlemin ne işe yaradığı açık değilse

If it is not clear which action is for which task

221 Hangi tuşa basınca ne olduğu açık değilse

If the function of the buttons are not clear

222 Kullanım sırasında alet beni bilgilendirmiyorsa

If device does not inform me during usage

223 Anlamsız bir sürü kısaltma kullanılıyorsa

If there are many meaningless abbreviations

224 Bana doğal gelmeyen bir kullanım şekli varsa

If style of use is not instinctive for me

225 Kullanımı mantığıma uygun değilse

If it does not fit my style of use

226 Bilindik terimler yerine yeni terimler kullanılıyorsa

If there new terms are used for common terms

227 Alet yaptıklarımı iptal etme şansı vermiyorsa

If device does not give me the opportunity to cancel what I do

228 Kullanım sırasında menüler arasında kayboluyorsam

If I get lost among menus during use

229 Alet hata yapmamı engelleyecek şekilde düşünülmemişse

If device does not prevent errors

230 Ciddi sonuçlara yol açabilecek hata yapma ihtimali varsa

If there is the possibility to make a mistake that may cause serious damage

231 Kullanım sırasında bir çok şeyi aklımda tutmam gerekiyorsa

If I have to recall many things while I use it

232 Kullanım sırasında gerekli bilgileri alet bana hatırlatmıyorsa

If device does not make me recall crucial information

233 En çok kullanacağım özelliklere ulaşmak çok zorsa

If it is hard to access frequenly used features

234 Menüleri kendi ihtiyaçlarıma göre düzenleyemiyorsam

If I cannot arrange menus according to my needs

235 Ekranlarda önemli bilgiler net olarak verilmiyorsa

If crucial information is not clearly displayed

236 Ekranda bir sürü gereksiz bilgi varsa

If there are lots of unnecessary information in the screen

237 Menülerde ihtiyacımdan çok daha fazla bilgi veriliyorsa.

If information provided in menus are more than I need

238 Alet karışık ekranlara sahipse

If device has complex screens

239 Hata uyarıları anlaşılmıyorsa

If error messages cannot be understood

240 Hata uyarıları beni çözüme yönlendirmiyorsa

If error messages does not lead me to solution

241 Hata oluştuğunda nedeni anlaşılamıyorsa

If I cannot understand the reason of an error

242 Hata uyarılarında anlaşılmaz sözcükler kullanılıyorsa

If there are incomprehensible words in error messages

APPENDIX G

RESULTS OF EXPERT REVIEW

		Form							Content							Composite
No	Item	R1	R2	R3	R4	R5	Av.	St.dev.	R1	R2	R3	R4	R5	Av.	St.dev.	Av.
1	Daha önce kullandığım tür bir alet değilse	7	4	3	7	3	4.80	2.05	8	7	6	9	6	7.20	1.30	6.00
2	Daha önceden kullanmadığım bir tür aletse	5	4	6	8	5	5.60	1.52	8	7	6	9	6	7.20	1.30	6.40
3	Daha önce aynı işe yarayan bir aleti kullanmadıysam	5	7	9	9	7	7.40	1.67	4	6	8	9	9	7.20	2.17	7.30
4	Daha önce karşılaşmadığım bir aletse	8	8	8	7	9	8.00	0.71	7	8	8	5	9	7.40	1.52	7.70
5	Daha önceden kullandığım aletlere benzemiyorsa	6	7	8	6	7	6.80	0.84	8	8	9	6	9	8.00	1.22	7.40
6	Kullanımı önceden bildiğim aletlere benzemiyorsa	8	6	4	5	3	5.20	1.92	5	7	8	7	6	6.60	1.14	5.90
7	Sık sık kullandığım aletlere benzemiyorsa	6	8	8	5	9	7.20	1.64	2	8	6	9	8	6.60	2.79	6.90
8	Diğer aletlerden bildiğim kullanım şeklini uygulayamıyorsam	4	7	7	5	4	5.40	1.52	9	8	8	5	9	7.80	1.64	6.60
9	Çok değişik özelliklere sahipse	2	1	5	7	5	4.00	2.45	1	4	8	6	7	5.20	2.77	4.60
10	Menüsü aynı tür aletlerin menüsüne benzemiyorsa	1	6	9	6	9	6.20	3.27	1	8	9	5	9	6.40	3.44	6.30
11	Diğer aletlere benzemiyorsa	1	5	6	4	5	4.20	1.92	1	4	3	4	8	4.00	2.55	4.10
12	Önceki aletlerden kazandığım tecrübeyi kullanamıyorsam	8	5	7	7	9	7.20	1.48	9	7	9	5	7	7.40	1.67	7.30
13	Daha önce benzer bir menüyle karşılaşmamışsam	1	7	8			5.33	3.79	2	8	9	*	9	7.00	3.37	6.17
14	Daha önce kullandığım aletlerden çok farklıysa	7	6	4	9	9	7.00	2.12	8	8	6	7	9	7.60	1.14	7.30
15	Bana yabancı bir aletse	8	6	6	5	7	6.40	1.14	7	6	7	3	7	6.00	1.73	6.20
16	Alıştığım bir markaya ait değilse	7	6	7	9	9	7.60	1.34	7	5	4	7	1	4.80	2.49	6.20
17	Aynı markaya ait başka alet kullanmamışsam	5	6	8	8	6	6.60	1.34	2	5	6	9	7	5.80	2.59	6.20
18	Herkes tarafından tercih edilen bir markaya ait değilse	1	7	7	9	1	5.00	3.74	1	8	6	5	1	4.20	3.11	4.60
19	Alıştığım bir aletin yeni modeli değilse	3	4	8	9	9	6.60	2.88	5	5	8	9	9	7.20	2.05	6.90
20	Daha önceki modelleriyle benzerlik taşımıyorsa	8	7	7	7	4	6.60	1.52	9	7	5	6	7	6.80	1.48	6.70
21	Daha önce alıştığım aletle arasında çok fark varsa	8	6	7	8	9	7.60	1.14	7	7	7	8	9	7.60	0.89	7.60
22	Aletin kullanımı yaygın değilse	5	8	8	6	4	6.20	1.79	2	8	7	6	6	5.80	2.28	6.00
23	Yeni teknolojiler içeriyorsa	1	8	7	9	7	6.40	3.13	2	8	8	7	8	6.60	2.61	6.50
24	Çok yeni bir aletse	1	6	5	6	3	4.20	2.17	2	7	3	5	5	4.40	1.95	4.30
25	Aletin ilk kullanıcılarındansam	2	6	8	6	8	6.00	2.45	1	7	8	3	3	4.40	2.97	5.20
26	Yaygın olmayan bir aletse	2	5	8	4	5	4.80	2.17	2	6	7	4	5	4.80	1.92	4.80
27	Kullanımı yaygın olmayan bir aletse	5	7	9	9	7	7.40	1.67	4	7	9	8	6	6.80	1.92	7.10
28	Alet ilgimi çekmemişse	5	6	7	6	7	6.20	0.84	4	8	8	4	5	5.80	2.05	6.00
29	Alet bana ilgi çekici gelmediyse	6	8	8	6	8	7.20	1.10	4	8	8	4	6	6.00	2.00	6.60

30	Çok ilgilenmediğim bir aletse	2	7	6	8	4	**5.40**	**2.41**	3	8	6	5	6	**5.60**	**1.82**	**5.50**
31	Alet ilgi alanıma girmiyorsa	2	7	9	9	6	**6.60**	**2.88**	8	8	9	9	7	**8.20**	**0.84**	**7.40**
32	Alete karşı ilgim fazla değilse	2	8	8	7	3	**5.60**	**2.88**	2	8	8	5	6	**5.80**	**2.49**	**5.70**
33	Sevdiğim tür bir alet değilse	3	6	7	4	3	**4.60**	**1.82**	1	6	7	4	8	**5.20**	**2.77**	**4.90**
34	Hoşlandığım bir alet değilse	4	6	8	4	3	**5.00**	**2.00**	2	6	8	4	5	**5.00**	**2.24**	**5.00**
35	Alete fazla ısınamadıysam	2	7	8	2	7	**5.20**	**2.95**	3	8	7	4	8	**6.00**	**2.35**	**5.60**
36	Aleti fazla sevmediysem	2	4	9	5	7	**5.40**	**2.70**	3	5	9	5	8	**6.00**	**2.45**	**5.70**
37	Aletten çok hoşlanmamışsam	1	4	8	5	7	**5.00**	**2.74**	3	5	8	4	8	**5.60**	**2.30**	**5.30**
38	Kullanmayı gerçekten istemiyorsam	7	4	6	5	9	**6.20**	**1.92**	9	6	9	5	9	**7.60**	**1.95**	**6.90**
39	Öğrenmeyi gerçekten istemiyorsam	7	5	6	4	9	**6.20**	**1.92**	7	7	7	4	9	**6.80**	**1.79**	**6.50**
40	Öğrenmekten zevk almıyorsam	3	5	9	4	1	**4.40**	**2.97**	6	7	9	3	1	**5.20**	**3.19**	**4.80**
41	Nasıl kullanıldığını çözmek hoşuma gitmiyorsa	8	6	7	9	9	**7.80**	**1.30**	3	8	9	9	6	**7.00**	**2.55**	**7.40**
42	Aleti kullanmak beni sıkıyorsa	8	5	9	6	9	**7.40**	**1.82**	7	5	9	3	9	**6.60**	**2.61**	**7.00**
43	Öğrenmekten çabuk sıkıldığım bir aletse	1	5	7	5	9	**5.40**	**2.97**	2	6	7	3	7	**5.00**	**2.35**	**5.20**
44	Alet bende merak uyandırmıyorsa	7	6	8	8	9	**7.60**	**1.14**	8	7	7	7	7	**7.20**	**0.45**	**7.40**
45	Alet bana itici geliyorsa	8	6	7	7	9	**7.40**	**1.14**	7	7	3	4	9	**6.00**	**2.45**	**6.70**
46	Severek aldığım bir alet değilse	8	6	9	8	6	**7.40**	**1.34**	6	7	9	6	7	**7.00**	**1.22**	**7.20**
47	Çok gerek görmediğim bir aletse	7	5	7	8	3	**6.00**	**2.00**	5	6	8	6	6	**6.20**	**1.10**	**6.10**
48	Özelliklerini çok fazla kullanmayacaksam	3	8	8	7	8	**6.80**	**2.17**	7	8	8	5	9	**7.40**	**1.52**	**7.10**
49	Fazla ihtiyaç duymadığım bir aletse	7	6	9	8	9	**7.80**	**1.30**	2	6	9	6	9	**6.40**	**2.88**	**7.10**
50	İhtiyaçlarımı karşılayacak bir alet değilse	5	6	5	6	7	**5.80**	**0.84**	3	6	8	5	8	**6.00**	**2.12**	**5.90**
51	İhtiyaçlarıma cevap verecek nitelikte değilse	2	6	6	8	6	**5.60**	**2.19**	4	6	3	4	8	**5.00**	**2.00**	**5.30**
52	Alet ihtiyaçtan alınmamışsa	1	8	4	7	9	**5.80**	**3.27**	2	8	6	5	7	**5.60**	**2.30**	**5.70**
53	Günlük hayatımı kolaylaştıracak bir alet değilse	5	6	4	7	9	**6.20**	**1.92**	3	6	2	4	7	**4.40**	**2.07**	**5.30**
54	İhtiyaçlarıma cevap vermiyorsa	2	8	9	5	6	**6.00**	**2.74**	3	8	8	3	8	**6.00**	**2.74**	**6.00**
55	İhtiyaçtan ötürü alınmış bir alet değilse	2	8	7	9	9	**7.00**	**2.92**	3	8	8	9	6	**6.80**	**2.39**	**6.90**
56	Günlük hayatta kullanabileceğim bir alet değilse	4	6	8	8	6	**6.40**	**1.67**	4	7	7	7	4	**5.80**	**1.64**	**6.10**
57	Kullanmayacağım özellikleri varsa	2	8	9	8	9	**7.20**	**2.95**	3	8	9	8	6	**6.80**	**2.39**	**7.00**
58	İşime yaramayacak özellikleri çoksa	5	8	7	7	9	**7.20**	**1.48**	4	8	9	8	6	**7.00**	**2.00**	**7.10**
59	İşime yaramayacak bir aletse	5	6	7	5	9	**6.40**	**1.67**	3	6	9	4	9	**6.20**	**2.77**	**6.30**
60	İşimi daha iyi yapmam için gerekli bir alet değilse	1	6	4	5	9	**5.00**	**2.92**	2	6	6	4	9	**5.40**	**2.61**	**5.20**
61	Yaptığım işleri daha iyi yapmamı sağlayacaksa	2	8	7	4	7	**5.60**	**2.51**	2	8	8	3	7	**5.60**	**2.88**	**5.60**
62	Özelliklerinin çoğu işime yaramıyorsa	2	8	8	8	7	**6.60**	**2.61**	3	9	9	8	8	**7.40**	**2.51**	**7.00**
63	Günlük hayatta sürekli kullanacağım bir alet değilse	5	7	3	8	9	**6.40**	**2.41**	7	8	7	7	6	**7.00**	**0.71**	**6.70**
64	Kullanmak zorunda olduğum bir alet değilse	5	7	7	6	8	**6.60**	**1.14**	8	6	8	5	9	**7.20**	**1.64**	**6.90**

65	Aleti kullanmam gerekli değilse	2	5	8	5	9	**5.80**	**2.77**	4	6	7	5	8	**6.00**	**1.58**	**5.90**
66	Sıkça kullanıdığım bir alet değilse	4	7	9	8	9	**7.40**	**2.07**	6	8	9	7	8	**7.60**	**1.14**	**7.50**
67	Sürekli kullanmam gerekmiyorsa	1	7	8	6	9	**6.20**	**3.11**	3	8	9	5	7	**6.40**	**2.41**	**6.30**
68	Aleti kullanmaya mecbur değilsem	1	5	7	7	9	**5.80**	**3.03**	3	6	8	5	9	**6.20**	**2.39**	**6.00**
69	Aleti kullanmam şart değilse	3	5	8	6	9	**6.20**	**2.39**	2	6	8	5	9	**6.00**	**2.74**	**6.10**
70	Basit bir alet değilse	6	6	8	5	9	**6.80**	**1.64**	6	8	6	4	6	**6.00**	**1.41**	**6.40**
71	Menüsü bana ters geliyorsa	1	1	6	3	5	**3.20**	**2.28**	2	1	8	2	7	**4.00**	**3.24**	**3.60**
72	Menü kullanımı kolay değilse	1	7	7	6	9	**6.00**	**3.00**	4	8	9	5	7	**6.60**	**2.07**	**6.30**
73	Menüsü açık - net değilse	1	7	9	7	9	**6.60**	**3.29**	4	8	9	6	9	**7.20**	**2.17**	**6.90**
74	Basit bir kullanımı yoksa	6	7	6	5	9	**6.60**	**1.52**	7	8	8	5	6	**6.80**	**1.30**	**6.70**
75	Nasıl kullanılacağı açık değilse	9	7	9	8	9	**8.40**	**0.89**	8	8	9	5	9	**7.80**	**1.64**	**8.10**
76	Kolay kullanılabilen bir alet değilse	9	7	7	4	7	**6.80**	**1.79**	7	7	8	3	6	**6.20**	**1.92**	**6.50**
77	Basit adımlarla istediğime ulaşmam mümkün değilse	4	8	5	4	7	**5.60**	**1.82**	8	8	7	2	9	**6.80**	**2.77**	**6.20**
78	İlk görüşte bana zor göründüyse	8	8	7	3	1	**5.40**	**3.21**	9	9	8	2	8	**7.20**	**2.95**	**6.30**
79	Kullanım açık değilse	7	7	6	4	6	**6.00**	**1.22**	8	8	8	4	9	**7.40**	**1.95**	**6.70**
80	Nasıl kullanılacağı net değilse	9	6	7	4	9	**7.00**	**2.12**	8	7	8	4	9	**7.20**	**1.92**	**7.10**
81	Kullanımı zor bir aletse	9	7	8	6	6	**7.20**	**1.30**	8	8	9	6	9	**8.00**	**1.22**	**7.60**
82	Aletin kullanımı karışıksa	9	7	9	6	9	**8.00**	**1.41**	8	8	9	5	9	**7.80**	**1.64**	**7.90**
83	Çok kullanılan özellikleri kolay bulunamıyorsa	2	8	9	8	9	**7.20**	**2.95**	2	9	9	6	9	**7.00**	**3.08**	**7.10**
84	Kullanım aşamaları akılda kalıcı değilse	2	8	5	9	7	**6.20**	**2.77**	7	9	8	8	9	**8.20**	**0.84**	**7.20**
85	Çalışma biçimini kavrayamadıysam	9	8	8	5	9	**7.80**	**1.64**	9	9	8	5	7	**7.60**	**1.67**	**7.70**
86	Tuşların ne işe yaradığı açık değilse	8	8	9	8	9	**8.40**	**0.55**	6	9	9	9	9	**8.40**	**1.34**	**8.40**
87	Hızlı bir şekilde istediğime ulaşamıyorsam	3	8	7	4	8	**6.00**	**2.35**	8	9	6	4	9	**7.20**	**2.17**	**6.60**
88	Kullanımı dolambaçlı olursa	3	5	6	3	9	**5.20**	**2.49**	6	7	7	3	9	**6.40**	**2.19**	**5.80**
89	Kullanım sırasında bir sürü aşamadan geçmek gerekiyorsa	8	3	9	4	8	**6.40**	**2.70**	7	7	9	3	9	**7.00**	**2.45**	**6.70**
90	Özelliklere hemen ulaşamıyorsam	7	3	5	5	9	**5.80**	**2.28**	8	7	7	4	9	**7.00**	**1.87**	**6.40**
91	Tuşların açıklamaları yoksa	2	7	4	6	9	**5.60**	**2.70**	3	8	8	6	9	**6.80**	**2.39**	**6.20**
92	Tuşların üstünde ne işe yaradıkları yazılı değilse	8	8	8	9	7	**8.00**	**0.71**	4	9	8	9	9	**7.80**	**2.17**	**7.90**
93	Tuşların üstündeki resimler belirgin değilse	7	8	7	8	7	**7.40**	**0.55**	6	9	8	6	8	**7.40**	**1.34**	**7.40**
94	Tuşların üstündeki açıklamalar diğer aletlerden farklıysa	7	6	8	8	1	**6.00**	**2.92**	7	9	8	6	1	**6.20**	**3.11**	**6.10**
95	Sık sık kılavuza başvurmam gerekiyorsa	9	8	9	6	9	**8.20**	**1.30**	9	9	9	4	8	**7.80**	**2.17**	**8.00**
96	İç güdülerime dayanarak çözmem mümkün değilse	4	5	7	3	1	**4.00**	**2.24**	8	8	6	2	9	**6.60**	**2.79**	**5.30**
97	Kullanım sırasında yönlendirmeler yoksa	2	8	7	7	6	**6.00**	**2.35**	8	9	8	5	9	**7.80**	**1.64**	**6.90**

98	Menülerde açıklamalar net değilse	1	8	8	7	8	**6.40**	**3.05**	4	9	8	5	8	**6.80**	**2.17**	**6.60**
99	Menülerde açıklayıcı bilgiler yoksa	1	8	7	6	6	**5.60**	**2.70**	3	9	8	4	9	**6.60**	**2.88**	**6.10**
100	Mantık yürüterek çözebileceğim bir alet değilse	7	8	8	4	7	**6.80**	**1.64**	8	8	8	4	9	**7.40**	**1.95**	**7.10**
101	İlk bakışta nasıl kullanılacağını anlayamadıysam	8	9	7	6	8	**7.60**	**1.14**	9	9	7	4	8	**7.40**	**2.07**	**7.50**
102	Temel özelliklerin nasıl kullanılacağı açık değilse	8	9	7	9	9	**8.40**	**0.89**	7	9	6	9	9	**8.00**	**1.41**	**8.20**
103	Kılavuza ihtiyaç duymadan alet kendi kendini anlatamıyorsa	9	9	4	6	7	**7.00**	**2.12**	8	9	7	4	9	**7.40**	**2.07**	**7.20**
104	Anlaşılmayan resimler-semboller varsa	9	8	7	7	7	**7.60**	**0.89**	7	8	8	6	8	**7.40**	**0.89**	**7.50**
105	Tuşların ne işe yaradığı anlaşılmıyorsa	9	8	9	8	8	**8.40**	**0.55**	6	8	9	9	9	**8.20**	**1.30**	**8.30**
106	Aletin üstünde belirsiz açıklamalar olursa	2	8	4	5	3	**4.40**	**2.30**	6	8	7	4	5	**6.00**	**1.58**	**5.20**
107	Kullanım şekli aletin üstünde gösterilmiyorsa	4	8	5	6	9	**6.40**	**2.07**	6	7	6	4	7	**6.00**	**1.22**	**6.20**
108	Aletin üzerindeki yazılar yönlendirici değilse	3	8	7	8	5	**6.20**	**2.17**	4	8	6	7	7	**6.40**	**1.52**	**6.30**
109	Aletin üstünde yönlendirici bilgiler yoksa	6	8	7	7	6	**6.80**	**0.84**	6	8	7	6	7	**6.80**	**0.84**	**6.80**
110	Kullanım sırasında yönlendirici bilgiler verilmiyorsa	4	8	9	8	4	**6.60**	**2.41**	7	9	8	6	7	**7.40**	**1.14**	**7.00**
111	Tuşlar birden fazla işe yarıyorsa	8	9	9	7	9	**8.40**	**0.89**	9	9	7	4	7	**7.20**	**2.05**	**7.80**
112	Çok fazla tuşu varsa	9	9	9	8	9	**8.80**	**0.45**	8	9	9	7	9	**8.40**	**0.89**	**8.60**
113	Menüsü çok karışıksa	1	7	9	6	9	**6.40**	**3.29**	5	8	9	5	9	**7.20**	**2.05**	**6.80**
114	Alet karmaşık bir yapıya sahipse	2	7	9	6	7	**6.20**	**2.59**	2	8	9	5	8	**6.40**	**2.88**	**6.30**
115	Menülerde çok fazla değişken varsa	1	6	4	7	1	**3.80**	**2.77**	9	7	6	5		**6.75**	**1.71**	**5.28**
116	Menüsü çok karışıksa	1		9	6	9	**6.25**	**3.77**	8		9	5	9	**7.75**	**1.89**	**7.00**
117	Alette çok menü varsa	1	6	9	6	3	**5.00**	**3.08**	8	7	9	6	3	**6.60**	**2.30**	**5.80**
118	Fazla alt menüsü varsa	1	8	7	6	9	**6.20**	**3.11**	9	8	6	6	9	**7.60**	**1.52**	**6.90**
119	Menüler çok karışık yapılmışsa	1	7	7	7	9	**6.20**	**3.03**	8	8	8	4	9	**7.40**	**1.95**	**6.80**
120	Menülerin içeriği çoksa	1	6	3	2	9	**4.20**	**3.27**	8	7	6	1	6	**5.60**	**2.70**	**4.90**
121	Menüler çok karmaşıksa	1	7	9	7	9	**6.60**	**3.29**	7	8	9	5	9	**7.60**	**1.67**	**7.10**
122	Alet çok karmaşık özelliklere sahipse	8	6	4	7	6	**6.20**	**1.48**	8	7	6	6	7	**6.80**	**0.84**	**6.50**
123	Alet karmaşıksa	3	6	8	6	9	**6.40**	**2.30**	8	7	9	5	9	**7.60**	**1.67**	**7.00**
124	Çok fazla özelliğe sahipse	2	7	9	9	9	**7.20**	**3.03**	7	8	9	8	9	**8.20**	**0.84**	**7.70**
125	Çok özelliği varsa	2	7	8		6	**5.75**	**2.63**	7	8	9	*	9	**8.25**	**0.96**	**7.00**
126	Çok amaçlı bir aletse	2	7	7	7	8	**6.20**	**2.39**	2	8	3	5	9	**5.40**	**3.05**	**5.80**
127	Özellikler iyi adlandırılmamışsa	1	6	7	8	9	**6.20**	**3.11**	2	8	9	6	8	**6.60**	**2.79**	**6.40**
128	Kullanılan teknik kelimeler anlaşılmaz olursa	7	7	8	9	9	**8.00**	**1.00**	8	8	7	9	9	**8.20**	**0.84**	**8.10**
129	Üstünde anlaşılmayan sözcükler varsa	7	5	6	8	6	**6.40**	**1.14**	7	8	8	7	7	**7.40**	**0.55**	**6.90**
130	Tuşların üstünde bilmediğim dilde yazılar varsa	8	7	7	9	9	**8.00**	**1.00**	7	9	9	4	9	**7.60**	**2.19**	**7.80**
131	Alette bilmediğim bir dil kullanılıyorsa	8	6	9	4	9	**7.20**	**2.17**	6	9	9	4	9	**7.40**	**2.30**	**7.30**
132	Alette kullanılan dil açık değilse	7	4	9	9	8	**7.40**	**2.07**	7	7	9	7	8	**7.60**	**0.89**	**7.50**

133	Satın aldığım yerde öğreten biri yoksa	8	7	5	5	7	**6.40**	**1.34**	7	9	7	3	6	**6.40**	**2.19**	**6.40**
134	Satılırken açıklayıcı bilgi verilmezse	7	8	4	8	9	**7.20**	**1.92**	7	9	7	6	6	**7.00**	**1.22**	**7.10**
135	Satan yer yardımcı olmazsa	5	6	3	4	5	**4.60**	**1.14**	3	8	7	3	5	**5.20**	**2.28**	**4.90**
136	Satıcı nasıl kullanacağımı göstermezse	4	8	8	6	9	**7.00**	**2.00**	3	9	8	3	6	**5.80**	**2.77**	**6.40**
137	Satış elemanı yardımcı olmazsa	4	8	4	4	5	**5.00**	**1.73**	3	9	7	3	5	**5.40**	**2.61**	**5.20**
138	Aleti kullananlardan bilgi alamıyorsam	5	6	4	3	7	**5.00**	**1.58**	4	9	6	1	7	**5.40**	**3.05**	**5.20**
139	Bilen kişilere sorma şansım yoksa	8	7	7	6	8	**7.20**	**0.84**	9	9	8	5	9	**8.00**	**1.73**	**7.60**
140	Bilen biri tarafından kullanım anlatılmazsa	8	7	8	7	8	**7.60**	**0.55**	9	9	9	6	9	**8.40**	**1.34**	**8.00**
141	Nasıl kullanıldığını özetleyebilecek biri yoksa	7	7	8	6	9	**7.40**	**1.14**	5	9	9	5	9	**7.40**	**2.19**	**7.40**
142	Kullanımı gösterecek biri yoksa	7	7	8	7	8	**7.40**	**0.55**	8	9	9	6	9	**8.20**	**1.30**	**7.80**
143	Aleti daha önce kullanmış bir arkadaşım yoksa	7	4	7	6	9	**6.60**	**1.82**	8	7	6	2	5	**5.60**	**2.30**	**6.10**
144	Zorlandığımda yardım alabileceğim biri yoksa	9	7	9	8	9	**8.40**	**0.89**	9	9	9	5	9	**8.20**	**1.79**	**8.30**
145	Kullanabilen birini gözlemleme şansım yoksa	8	8	9	8	9	**8.40**	**0.55**	9	9	8	8	9	**8.60**	**0.55**	**8.50**
146	Aleti bana öğretecek bir tanıdık yoksa	9	5	9	6	8	**7.40**	**1.82**	9	8	7	4	8	**7.20**	**1.92**	**7.30**
147	Bilen birinden yardım alamıyorsam	9	7	9	6	9	**8.00**	**1.41**	8	8	8	5	9	**7.60**	**1.52**	**7.80**
148	Öğrenmemi destekleyecek biri yoksa	7	5	6	7	5	**6.00**	**1.00**	8	8	7	2	7	**6.40**	**2.51**	**6.20**
149	Daha önce kullananlardan destek alamıyorsam	8	6	7	3	4	**5.60**	**2.07**	8	7	7	1	6	**5.80**	**2.77**	**5.70**
150	Daha önce kullananlara danışma fırsatım yoksa	8	7	8	6	8	**7.40**	**0.89**	8	9	7	5	7	**7.20**	**1.48**	**7.30**
151	Kullanımı bilen biri uygulamalı olarak anlatmazsa	2	8	8	7	8	**6.60**	**2.61**	7	9	9	6	9	**8.00**	**1.41**	**7.30**
152	Yardım alabileceğim kimse yoksa	6	6	9	6	9	**7.20**	**1.64**	7	7	9	6	9	**7.60**	**1.34**	**7.40**
153	Çevremde kullanan başka insanlar yoksa	5	7	7	5	9	**6.60**	**1.67**	7	7	7	4	8	**6.60**	**1.52**	**6.60**
154	Takıldığım zaman yardım edecek kimse yoksa	7	8	9	4	9	**7.40**	**2.07**	8	9	9	3	9	**7.60**	**2.61**	**7.50**
155	Kullanımı gösterecek kişiler yoksa	7	7	9	7	8	**7.60**	**0.89**	8	8	9	6	9	**8.00**	**1.22**	**7.80**
156	Bilgi alabileceğim kimse yoksa	5	7	7	6	6	**6.20**	**0.84**	7	8	8	6	9	**7.60**	**1.14**	**6.90**
157	Çevremde aleti bilen biri yoksa	8	5	7	3	9	**6.40**	**2.41**	7	7	7	2	9	**6.40**	**2.61**	**6.40**
158	Yönlendirecek biri yoksa	2	6	7	5	6	**5.20**	**1.92**	5	7	8	5	7	**6.40**	**1.34**	**5.80**
159	Detaylı şekilde anlatacak biri yoksa	5	7	8	6	9	**7.00**	**1.58**	7	8	9	6	9	**7.80**	**1.30**	**7.40**
160	Kılavuzu yoksa	8	8	9	8	9	**8.40**	**0.55**	8	8	9	6	9	**8.00**	**1.22**	**8.20**
161	İyi bir yardım menüsüne sahip değilse	2	9	7	6	9	**6.60**	**2.88**	7	9	6	4	9	**7.00**	**2.12**	**6.80**
162	Kılavuzda kullanımı kısaca anlatan bir bölüm yoksa	8	9	8	6	9	**8.00**	**1.22**	5	9	7	7	9	**7.40**	**1.67**	**7.70**
163	Alet içinde kullanımı öğreten bir bölüm yoksa	2	6	6	7	3	**4.80**	**2.17**	3	7	3	7	6	**5.20**	**2.05**	**5.00**
164	Kılavuz anlaşılamıyorsa	8	8	8	6	9	**7.80**	**1.10**	7	8	7	6	9	**7.40**	**1.14**	**7.60**
165	Kılavuzda verilen bilgiler net değilse	6	8	9	7	8	**7.60**	**1.14**	7	8	9	6	8	**7.60**	**1.14**	**7.60**
166	Kılavuz iyi değilse	4	6	7	4	7	**5.60**	**1.52**	8	6	8	2	7	**6.20**	**2.49**	**5.90**
167	Kılavuz yetersizse	4	6	9	4	6	**5.80**	**2.05**	8	7	9	3	7	**6.80**	**2.28**	**6.30**

	düşünüyorsam															
200	Kullanmaya çekindiğim bir aletse	7	7	8	2	3	**5.40**	**2.70**	8	7	7	1	1	**4.80**	**3.49**	**5.10**
201	Yanlış kullanıldığında başa dönmek zorsa	3	7	3	6	5	**4.80**	**1.79**	8	8	7	5	7	**7.00**	**1.22**	**5.90**
202	Alette kullanılan kısaltmaların ne anlama geldiğini bilmiyorsam	6	8	8	6	9	**7.40**	**1.34**	7	8	8	1	8	**6.40**	**3.05**	**6.90**
203	Kulllanılan terimlerin ne anlama geldiğini bilmiyorsam	6	8	7	6	7	**6.80**	**0.84**	8	8	8	1	8	**6.60**	**3.13**	**6.70**
204	Çok fazla özel terim kullanılıyorsa	5	8	6	7	6	**6.40**	**1.14**	4	8	7	5	8	**6.40**	**1.82**	**6.40**
205	Çok fazla kısaltma kullanılıyorsa	5	8	8	5	6	**6.40**	**1.52**	4	8	8	2	7	**5.80**	**2.68**	**6.10**
206	Gerekli bilgiye sahip değilsem	8	5	6	4	6	**5.80**	**1.48**	8	7	4	1	8	**5.60**	**3.05**	**5.70**
207	Daha önceden alet hakkında bilgim yoksa	7	7	7	5	8	**6.80**	**1.10**	7	7	7	2	7	**6.00**	**2.24**	**6.40**
208	Alet bilgi birikimim dışında bilgi gerektiriyorsa	8	7	6	5	8	**6.80**	**1.30**	8	7	8	2	9	**6.80**	**2.77**	**6.80**
209	Çok karışık bilgi içeriyorsa	8	5	7	5	1	**5.20**	**2.68**	8	7	6	3	1	**5.00**	**2.92**	**5.10**
210	İyi düşünülerek yapılmamış bir alet değilse	5	5	1	3	7	**4.20**	**2.28**	1	6	3	3	5	**3.60**	**1.95**	**3.90**
211	Menüsü kötü yapılmışsa	1	7	7	3	7	**5.00**	**2.83**	8	7	8	2	5	**6.00**	**2.55**	**5.50**
212	Menüleri kolay kullanıma göre yapılmadıysa	1	7	2	5	7	**4.40**	**2.79**	2	8	5	5	5	**5.00**	**2.12**	**4.70**
213	Kullanım kolaylığı düşünülmeden yapılmış bir aletse	8	6	5	4	7	**6.00**	**1.58**	8	8	7	2	6	**6.20**	**2.49**	**6.10**
214	Bilmediğim bir konuyla ilgiliyse	3	4	7	4	7	**5.00**	**1.87**	8	6	9	2	7	**6.40**	**2.70**	**5.70**
215	Zor kontrol edilen bir aletse	7	6	6	4	7	**6.00**	**1.22**	8	8	8	2	9	**7.00**	**2.83**	**6.50**
216	Aletle yapılabilecek çok şey varsa	2	4	4	1	9	**4.00**	**3.08**	4	6	7	1	9	**5.40**	**3.05**	**4.70**
217	Kullanmadan önce bir sürü ayar yapmak gerekiyorsa	4	6	9	5	9	**6.60**	**2.30**	7	8	8	5	9	**7.40**	**1.52**	**7.00**
218	İlk kez açıldığında ayarlanması gereken çok şey varsa	7	6	8	4	9	**6.80**	**1.92**	8	8	8	5	7	**7.20**	**1.30**	**7.00**
219	Yaptıklarımın doğru mu yanlış mı olduğunu anlamakta zorlanıyorsam	7	7	9	2	6	**6.20**	**2.59**	9	9	9	2	9	**7.60**	**3.13**	**6.90**
220	Hangi işlemin ne işe yaradığı açık değilse	8	7	5	6	5	**6.20**	**1.30**	9	9	7	7	1	**6.60**	**3.29**	**6.40**
221	Hangi tuşa basınca ne olduğu açık değilse	2	7	8	7	9	**6.60**	**2.70**	8	9	8	7	9	**8.20**	**0.84**	**7.40**
222	Kullanım sırasında alet beni bilgilendirmiyorsa	2	6	3	6	6	**4.60**	**1.95**	7	8	4	4	8	**6.20**	**2.05**	**5.40**
223	Anlamsız bir sürü kısaltma kullanılıyorsa	4	6	8	1	7	**5.20**	**2.77**	4	9	6	1	8	**5.60**	**3.21**	**5.40**
224	Bana doğal gelmeyen bir kullanım şekli varsa	8	7	7	3	7	**6.40**	**1.95**	8	9	8	3	9	**7.40**	**2.51**	**6.90**
225	Kullanımı mantığıma uygun değilse	8	7	8	3	9	**7.00**	**2.35**	9	8	8	2	9	**7.20**	**2.95**	**7.10**
226	Bilindik terimler yerine yeni terimler kullanılıyorsa	6	7	8	4	9	**6.80**	**1.92**	4	9	4	3	9	**5.80**	**2.95**	**6.30**
227	Alet yaptıklarımı iptal etme şansı vermiyorsa	5	8	7	5	9	**6.80**	**1.79**	7	9	9	3	9	**7.40**	**2.61**	**7.10**
228	Kullanım sırasında menüler arasında kayboluyorsam	2	9	9	3	9	**6.40**	**3.58**	8	9	9	1	9	**7.20**	**3.49**	**6.80**

229	Alet hata yapmamı engelleyecek şekilde düşünülmemişse	7	9	6	4	9	**7.00**	**2.12**	9	9	7	2	7	**6.80**	**2.86**	**6.90**
230	Ciddi sonuçlara yol açabilecek hata yapma ihtimali varsa	9	7	7	3	9	**7.00**	**2.45**	9	9	8	2	9	**7.40**	**3.05**	**7.20**
231	Kullanım sırasında bir çok şeyi aklımda tutmam gerekiyorsa	9	6	9	7	6	**7.40**	**1.52**	8	9	9	7	9	**8.40**	**0.89**	**7.90**
232	Kullanım sırasında gerekli bilgileri alet bana hatırlatmıyorsa	5	7	7	8	5	**6.40**	**1.34**	8	9	8	7	9	**8.20**	**0.84**	**7.30**
233	En çok kullanacağım özelliklere ulaşmak çok zorsa	6	7	7	8	9	**7.40**	**1.14**	8	9	7	6	9	**7.80**	**1.30**	**7.60**
234	Menüleri kendi ihtiyaçlarıma göre düzenleyemiyorsam	2	6	4	8	9	**5.80**	**2.86**	8	8	3	8	7	**6.80**	**2.17**	**6.30**
235	Ekranlarda önemli bilgiler net olarak verilmiyorsa	8	7	7	9	7	**7.60**	**0.89**	9	9	7	8	9	**8.40**	**0.89**	**8.00**
236	Ekranda bir sürü gereksiz bilgi varsa	9	7	8	5	7	**7.20**	**1.48**	9	8	7	4	9	**7.40**	**2.07**	**7.30**
237	Menülerde ihtiyacımdan çok daha fazla bilgi veriliyorsa.	2	7	7	5	5	**5.20**	**2.05**	8	8	8	4	7	**7.00**	**1.73**	**6.10**
238	Alet karışık ekranlara sahipse	7	5	8	7	9	**7.20**	**1.48**	7	9	9	4	9	**7.60**	**2.19**	**7.40**
239	Hata uyarıları anlaşılmıyorsa	7	7	9	9	9	**8.20**	**1.10**	5	9	9	5	9	**7.40**	**2.19**	**7.80**
240	Hata uyarıları beni çözüme yönlendirmiyorsa	7	7	8	7	9	**7.60**	**0.89**	8	9	8	5	9	**7.80**	**1.64**	**7.70**
241	Hata oluştuğunda nedeni anlaşılamıyorsa	7	7	9	9	9	**8.20**	**1.10**	8	9	9	9	9	**8.80**	**0.45**	**8.50**
242	Hata uyarılarında anlaşılmaz sözcükler kullanılıyorsa	7	7	7	4	9	**6.80**	**1.79**	6	9	8	3	7	**6.60**	**2.30**	**6.70**

APPENDIX H

CONSENT FORM

Gönülü Katılım Formu

Bu çalışma, Orta Doğu Teknik Üniversitesi Doktora Öğrencilerinden Ali Emre Berkman tarafından yürütülen "Demografik Özelliklerin Elektronik Alet Öğrenme Özgüvenine Etkisi" başlıklı araştırma kapsamında yürütülmektedir. Çalışmanın amacı, katılımcıların elektronik alet öğrenimle ilgili eğilimlerini saptamak ve yaş, cinsiyet, eğitim durumu gibi faktörlerin etkisini incelemektir. Katılımcılar çalışmaya gönüllü olarak katılmalıdırlar. Anket tamamen bilimsel amaçlı olarak değerlendirilecek ve hiçbir şekilde ticari amaçla kullanılmayacaktır. Verdiğiniz cevaplarla ve kimlik bilgileriniz eşleştirilmeyecek ve kimliğiniz her zaman gizli tutulacaktır.

Ankette yer alan sorular size rahatsızlık vermeyecek şekilde düzenlenmiştir. Soruları cevaplarken kendinizi herhangi bir nedenden ötürü rahatsız hissederseniz çalışmayı neden göstermeksizin yarıda bırakabilirsiniz. Böyle bir durumda anketi uygulayan kişiye, anketi tamamlamadığınızı söylemek yeterli olacaktır. Çalışmamıza verdiğiniz destekten ötürü size şimdiden çok teşekkür ederiz.

Çalışma hakkında daha ayrıntılı bilgi için:

Ali Berkman (Tel: 210 22 14 - 427 04 02; E-posta: ali.berkman@gmail.com) ile iletişim kurabilirsiniz.

Bu çalışmaya tamamen gönüllü olarak katılıyorum ve istediğim zaman yarıda kesip çıkabileceğimi biliyorum. Verdiğim bilgilerin bilimsel amaçlı yayımlarda kullanılmasını kabul ediyorum. (Formu doldurup imzaladıktan sonra uygulayıcıya geri veriniz).

İsim Soyad	Tarih	İmza
	----/----/-----	

APPENDIX I

GISE-S FORM: ITEM TRYOUT PHASE (SAMPLE)

Elektronik Aletlerin Öğrenimi Anketi

Cinsiyet:	Kadın **O** Erkek **O**
Eğitim durumu:	Yok **O** İlkokul **O** Ortaokul **O** Lise **O** Üniversite **O** Yüksek Lisans **O**
Yaş:	_____

Bu ankette **elektronik alet** veya **alet** olarak tanımlanan ürünler

Cep telefonları;
Dijital fotoğraf makinaları, video kameralar;
Dijital ekranlı fırın, çamaşır makinası, bulaşık makinası gibi ürünler;
MP3 çalarlar;
Elektronik ajandalar, sözlükler;
DVD/VCD oynatıcılar ve kaydediciler, uydu alıcılar;
Müzik setleri;
Televizyonlar;
Elektronik oyunlar;
Bankamatik cihazları...

...gibi **ekran ve tuşlar aracılığıyla** kullanılan ürünlerdir. Lütfen soruları bu tanıma giren ürünleri düşünerek cevaplayınız.

Anket formunda, sorularda anlatılan durumları düşününüz. **Anlatılan durum geçerli olsaydı bir elektronik aleti öğrenme konusunda kendinize ne kadar güvenirdiniz?** Değerlendirmeyi bir elektronik aleti öğrenebilme konusunda kendinize duyduğunuz güvene göre yapınız. Bu amaçla aşağıda verilen puanlama şeklini kullanınız.

Soruları 0-10 arasında puan vererek yanıtlayınız. Kendinize güvenmiyorsanız düşük, güveniyorsanız yüksek puanlar veriniz.

Puanlama

0 1 2 3 4 5 6 7 8 9 10

Aleti öğrenebileceğime kesinlikle **güvenmiyorum** — Aleti öğrenebileceğime kesinlikle **güveniyorum**

1/6

Puanlama

	0 1 2 3 4 5 6 7 8 9 10	
Aleti öğrenebileceğime kesinlikle **güven<u>miyorum</u>**		Aleti öğrenebileceğime kesinlikle **<u>güveniyorum</u>**

1.	Daha önce aynı işe yarayan bir aleti kullanmadıysam	Puan(0-10): ____
2.	Daha önce karşılaşmadığım bir aletse	Puan(0-10): ____
3.	Daha önceden kullandığım aletlere benzemiyorsa	Puan(0-10): ____
4.	Önceki aletlerden kazandığım tecrübeyi kullanamıyorsam	Puan(0-10): ____
5.	Daha önce kullandığım aletlerden çok farklıysa	Puan(0-10): ____
6.	Diğer aletlerden alıştığım kullanım şeklini uygulayamıyorsam	Puan(0-10): ____
7.	Daha önce alıştığım aletlerle arasında çok fark varsa	Puan(0-10): ____
8.	Kullanımı yaygın olmayan bir aletse	Puan(0-10): ____
9.	Daha önceki modelleriyle benzerlik taşımıyorsa	Puan(0-10): ____
10.	Kullanmaya alışık olmadığım teknolojiler içeriyorsa	Puan(0-10): ____
11.	İlgi alanıma girmiyorsa	Puan(0-10): ____
12.	Bana ilgi çekici gelmediyse	Puan(0-10): ____
13.	Nasıl kullanıldığını çözmek hoşuma gitmiyorsa	Puan(0-10): ____
14.	Severek aldığım bir alet değilse	Puan(0-10): ____
15.	Bende merak uyandırmıyorsa	Puan(0-10): ____
16.	Kullanmaktan sıkılıyorsam	Puan(0-10): ____
17.	Kullanmayı gerçekten istemiyorsam	Puan(0-10): ____
18.	Kullanmayacağım özellikleri varsa	Puan(0-10): ____
19.	İşime yaramayacak özellikleri çoksa	Puan(0-10): ____
20.	İhtiyaçtan ötürü alınmış bir alet değilse	Puan(0-10): ____
21.	Tüm özelliklerini kullanmayacaksam	Puan(0-10): ____
22.	Fazla ihtiyaç duymadığım bir aletse	Puan(0-10): ____
23.	İşime yarayacak bir alet değilse	Puan(0-10): ____
24.	Yaptığım işleri daha iyi yapmamı sağlamayacaksa	Puan(0-10): ____
25.	Özelliklerinin çoğu işime yaramıyorsa	Puan(0-10): ____
26.	Aleti kullanmam zorunlu değilse	Puan(0-10): ____
27.	Sıkça kullanacağım bir alet değilse	Puan(0-10): ____

Puanlama		
	0 1 2 3 4 5 6 7 8 9 10	
Aleti öğrenebileceğime kesinlikle **güvenmiyorum**		Aleti öğrenebileceğime kesinlikle **güveniyorum**

28.	Kullanmak zorunda olduğum bir alet değilse	Puan(0-10): ____
29.	Nasıl kullanılacağı açık değilse	Puan(0-10): ____
30.	Kullanımı zor geliyorsa	Puan(0-10): ____
31.	Aletin kullanımı karışıksa	Puan(0-10): ____
32.	Çok kullanılan özelliklerini bulmak kolay değilse	Puan(0-10): ____
33.	Kullanımı akılda kalıcı değilse	Puan(0-10): ____
34.	Çalışma biçimini kavrayamadıysam	Puan(0-10): ____
35.	Hızlı bir şekilde istediğime ulaşamıyorsam	Puan(0-10): ____
36.	Tuşların üstünde ne işe yaradıkları yazmıyorsa	Puan(0-10): ____
37.	Tuşların üstündeki resimler belirgin değilse	Puan(0-10): ____
38.	Sık sık kılavuza başvurmam gerekiyorsa	Puan(0-10): ____
39.	Mantık yürüterek çözebileceğim bir alet değilse	Puan(0-10): ____
40.	Temel özelliklerin nasıl kullanılacağı açık değilse	Puan(0-10): ____
41.	İlk bakışta nasıl kullanılacağını anlayamıyorsam	Puan(0-10): ____
42.	Kılavuza ihtiyaç duymadan alet kendi kendini anlatamıyorsa	Puan(0-10): ____
43.	Kendi kendime çözmem mümkün değilse	Puan(0-10): ____
44.	Tuşlar birden fazla işe yarıyorsa	Puan(0-10): ____
45.	Çok fazla tuşu varsa	Puan(0-10): ____
46.	Menüsü çok karışıksa	Puan(0-10): ____
47.	Çok karmaşık özelliklere sahipse	Puan(0-10): ____
48.	Alet karmaşıksa	Puan(0-10): ____
49.	Çok fazla özelliğe sahipse	Puan(0-10): ____
50.	Kullanılan teknik kelimeler anlaşılmıyorsa	Puan(0-10): ____
51.	Tuşların üstünde bilmediğim dilde yazılar varsa	Puan(0-10): ____
52.	Alette bilmediğim bir dil kullanılıyorsa	Puan(0-10): ____
53.	Kullanılan dil açık değilse	Puan(0-10): ____
54.	Satın alırken açıklayıcı bilgi verilmezse	Puan(0-10): ____
55.	Satıcı nasıl kullanacağımı göstermezse	Puan(0-10): ____

3/6

Puanlama		
	0 1 2 3 4 5 6 7 8 9 10	
Aleti öğrenebileceğime kesinlikle **güvenmiyorum**		Aleti öğrenebileceğime kesinlikle **güveniyorum**

56.	Bilen kişilere sorma şansım yoksa	Puan(0-10): ____
57.	Bilen biri tarafından kullanım anlatılmazsa	Puan(0-10): ____
58.	Kullanımı gösterecek biri yoksa	Puan(0-10): ____
59.	Zorlandığımda yardım alabileceğim biri yoksa	Puan(0-10): ____
60.	Kullanabilen birini gözlemleme şansım yoksa	Puan(0-10): ____
61.	Yardım alabileceğim kimse yoksa	Puan(0-10): ____
62.	Takıldığım zaman yardım edecek kimse yoksa	Puan(0-10): ____
63.	Kılavuzu yoksa	Puan(0-10): ____
64.	İyi bir yardım menüsüne sahip değilse	Puan(0-10): ____
65.	Kılavuzda verilen bilgiler net değilse	Puan(0-10): ____
66.	Kılavuz yeterince açıklayıcı değilse	Puan(0-10): ____
67.	Kılavuzda kullanım adım adım anlatılmıyorsa	Puan(0-10): ____
68.	Kılavuz anlaşılamıyorsa	Puan(0-10): ____
69.	Kullanım kılavuzu yeterince anlaşılır değilse	Puan(0-10): ____
70.	Kullanım kılavuzu açıklayıcı değilse	Puan(0-10): ____
71.	Kullanım kılavuzunda yalın bir dil yoksa	Puan(0-10): ____
72.	Kullanım kılavuzunda günlük dilde kullanılmayan sözcükler bulunuyorsa	Puan(0-10): ____
73.	Kılavuzda teknik terimler kullanılıyorsa	Puan(0-10): ____
74.	Teknik servisten telefonla yardım almak mümkün değilse	Puan(0-10): ____
75.	İstediğim kadar deneme yapma şansım yoksa	Puan(0-10): ____
76.	Kullanabilmek için önce sayfalarca kılavuz okumam gerekiyorsa	Puan(0-10): ____
77.	Bir kaç kez kullandığımda hala sorun yaşıyorsam	Puan(0-10): ____
78.	İlk kullanımda sorun yaşarsam	Puan(0-10): ____
79.	Kullanırken çok hata yapıyorsam	Puan(0-10): ____
80.	Aleti sıkça kullanma fırsatı bulamıyorsam	Puan(0-10): ____
81.	Yanımda zaten o aleti kullanmayı üstlenmiş biri varsa	Puan(0-10): ____
82.	Denerken aletin bozulma ihtimali varsa	Puan(0-10): ____
83.	Yanlış yaptığımda geri dönüş yoksa	Puan(0-10): ____
84.	Çabuk arızalanacak bir alet olduğunu düşünüyorsam	Puan(0-10): ____

4/6

Puanlama											
	0	1	2	3	4	5	6	7	8	9	10
Aleti öğrenebileceğime kesinlikle **güven<u>miyorum</u>**											Aleti öğrenebileceğime kesinlikle **<u>güveniyorum</u>**

85.	Alette kullanılan kısaltmaların ne anlama geldiğini bilmiyorsam	Puan(0-10): ____
86.	Daha önceden alet hakkında bilgim yoksa	Puan(0-10): ____
87.	Alet bilgi birikimim dışında bilgi gerektiriyorsa	Puan(0-10): ____
88.	Zor kontrol edilen bir aletse	Puan(0-10): ____
89.	Kullanmadan önce bir sürü ayar yapmak gerekiyorsa	Puan(0-10): ____
90.	İlk kez açıldığında ayarlanması gereken çok şey varsa	Puan(0-10): ____
91.	Yaptıklarımın doğru mu yanlış mı olduğunu anlamakta zorlanıyorsam	Puan(0-10): ____
92.	Hangi tuşa basınca ne olduğu açık değilse	Puan(0-10): ____
93.	Kullanımı mantığıma uygun değilse	Puan(0-10): ____
94.	Alet yaptıklarımı iptal etme şansı vermiyorsa	Puan(0-10): ____
95.	Ciddi sonuçlara yol açabilecek hata yapma ihtimali varsa	Puan(0-10): ____
96.	Kullanım sırasında bir çok şeyi aklımda tutmam gerekiyorsa	Puan(0-10): ____
97.	Kullanım sırasında gerekli bilgileri alet bana hatırlatmıyorsa	Puan(0-10): ____
98.	Ekranda önemli bilgiler net olarak verilmiyorsa	Puan(0-10): ____
99.	Menülerde ihtiyacımdan çok daha fazla bilgi veriliyorsa	Puan(0-10): ____
100.	Hata uyarıları anlaşılmıyorsa	Puan(0-10): ____
101.	Hata uyarıları beni çözüme yönlendirmiyorsa	Puan(0-10): ____
102.	Hata oluştuğunda nedeni anlaşılamıyorsa	Puan(0-10): ____
103.	Ekranda bir sürü gereksiz bilgi varsa	Puan(0-10): ____
104.	Alet karışık ekranlara sahipse	Puan(0-10): ____

Lütfen kullandığınız elektronik aletleri işaretleyiniz. Şu an artık kullanmıyor olsanız da, hayatınızın bir döneminde, evde, işte, arkadaşlarınızın veya akrabalarınızın evinde kullandığınız tüm ürünleri işaretleyiniz.

Cep telefonu	**O**	Dijital fotoğraf makinası	**O**	Video kamera	**O**	Ekranlı fırın	**O**
Ekranlı buzdolabı	**O**	Ekranlı çamaşır makinası	**O**	Ekranlı bulaşık makinası	**O**	Ekranlı kurutma makinası	**O**
MP3 çalar	**O**	Elektronik ajanda	**O**	Elektronik sözlük	**O**	DVD/VCD çalar	**O**
DVD/VCD kaydedici	**O**	Müzik seti	**O**	Televizyon	**O**	Elektronik oyun	**O**
Uydu alıcı	**O**	Bankamatik	**O**				

Çalışmaya katıldığınız için çok teşekkür ederiz…

5/6

APPENDIX J

GISE-S FORM: MAJOR DATA COLLECTION PHASE (SAMPLE)

Elektronik Aletlerin Öğrenimi Anketi

Cinsiyet: Kadın **O** Erkek **O**

Eğitim durumu: Yok **O** İlkokul **O** Ortaokul **O** Lise **O** Üniversite **O** Yüksek Lisans **O**

Yaş: _____

Bu ankette **elektronik alet** veya **alet** olarak tanımlanan ürünler

Cep telefonları;
Dijital fotoğraf makinaları, video kameralar;
Dijital ekranlı fırın, çamaşır makinası, bulaşık makinası gibi ürünler;
MP3 çalarlar;
Elektronik ajandalar, sözlükler;
DVD/VCD oynatıcılar ve kaydediciler, uydu alıcılar;
Müzik setleri;
Televizyonlar;
Elektronik oyunlar;
Bankamatik cihazları...

...gibi **<u>ekran ve tuşlar aracılığıyla</u>** kullanılan ürünlerdir. Lütfen soruları bu tanıma giren ürünleri düşünerek cevaplayınız.

Anket formunda, sorularda anlatılan durumları düşününüz. **Anlatılan durum geçerli olsaydı bir elektronik aleti öğrenme konusunda kendinize ne kadar güvenirdiniz?** Değerlendirmeyi bir elektronik aleti öğrenebilme konusunda kendinize duyduğunuz güvene göre yapınız. Bu amaçla aşağıda verilen puanlama şeklini kullanınız.

Soruları 0-10 arasında puan vererek yanıtlayınız. Kendinize güvenmiyorsanız düşük, güveniyorsanız yüksek puanlar veriniz.

Puanlama

0	1	2	3	4	5	6	7	8	9	10
Aleti öğrenebileceğime kesinlikle **güven<u>miyorum</u>**										Aleti öğrenebileceğime kesinlikle **<u>güveniyorum</u>**

1/5

Puanlama

	0	1	2	3	4	5	6	7	8	9	10	
Aleti öğrenebileceğime kesinlikle **güven<u>miyorum</u>**												Aleti öğrenebileceğime kesinlikle **<u>güveniyorum</u>**

1.	Daha önce aynı işe yarayan bir aleti kullanmadıysam	Puan(0-10): ____
2.	Daha önce karşılaşmadığım bir aletse	Puan(0-10): ____
3.	Daha önceden kullandığım aletlere benzemiyorsa	Puan(0-10): ____
4.	Önceki aletlerden kazandığım tecrübeyi kullanamıyorsam	Puan(0-10): ____
5.	Daha önce kullandığım aletlerden çok farklıysa	Puan(0-10): ____
6.	Diğer aletlerden alıştığım kullanım şeklini uygulayamıyorsam	Puan(0-10): ____
7.	Daha önce alıştığım aletlerle arasında çok fark varsa	Puan(0-10): ____
8.	Kullanımı yaygın olmayan bir aletse	Puan(0-10): ____
9.	Daha önceki modelleriyle benzerlik taşımıyorsa	Puan(0-10): ____
10.	Kullanmaya alışık olmadığım teknolojiler içeriyorsa	Puan(0-10): ____
11.	İlgi alanıma girmiyorsa	Puan(0-10): ____
12.	Bana ilgi çekici gelmediyse	Puan(0-10): ____
13.	Severek aldığım bir alet değilse	Puan(0-10): ____
14.	Kullanmaktan sıkılıyorsam	Puan(0-10): ____
15.	Kullanmayacağım özellikleri varsa	Puan(0-10): ____
16.	İşime yaramayacak özellikleri çoksa	Puan(0-10): ____
17.	Tüm özelliklerini kullanmayacaksam	Puan(0-10): ____
18.	Fazla ihtiyaç duymadığım bir aletse	Puan(0-10): ____
19.	İşime yarayacak bir alet değilse	Puan(0-10): ____
20.	Yaptığım işleri daha iyi yapmamı sağlamayacaksa	Puan(0-10): ____
21.	Sıkça kullanacağım bir alet değilse	Puan(0-10): ____
22.	Nasıl kullanılacağı açık değilse	Puan(0-10): ____
23.	Kullanımı zor geliyorsa	Puan(0-10): ____
24.	Aletin kullanımı karışıksa	Puan(0-10): ____
25.	Çok kullanılan özelliklerini bulmak kolay değilse	Puan(0-10): ____
26.	Kullanımı akılda kalıcı değilse	Puan(0-10): ____
27.	Çalışma biçimini kavrayamadıysam	Puan(0-10): ____

2/5

Puanlama		
	0 1 2 3 4 5 6 7 8 9 10	
Aleti öğrenebileceğime kesinlikle **güvenmiyorum**		Aleti öğrenebileceğime kesinlikle **güveniyorum**

28.	Hızlı bir şekilde istediğime ulaşamıyorsam	Puan(0-10): ____
29.	Tuşların üstünde ne işe yaradıkları yazmıyorsa	Puan(0-10): ____
30.	Tuşların üstündeki resimler belirgin değilse	Puan(0-10): ____
31.	Sık sık kılavuza başvurmam gerekiyorsa	Puan(0-10): ____
32.	Mantık yürüterek çözebileceğim bir alet değilse	Puan(0-10): ____
33.	Temel özelliklerin nasıl kullanılacağı açık değilse	Puan(0-10): ____
34.	Kendi kendime çözmem mümkün değilse	Puan(0-10): ____
35.	Tuşlar birden fazla işe yarıyorsa	Puan(0-10): ____
36.	Çok fazla tuşu varsa	Puan(0-10): ____
37.	Menüsü çok karışıksa	Puan(0-10): ____
38.	Çok karmaşık özelliklere sahipse	Puan(0-10): ____
39.	Alet karmaşıksa	Puan(0-10): ____
40.	Çok fazla özelliğe sahipse	Puan(0-10): ____
41.	Kullanılan teknik kelimeler anlaşılmıyorsa	Puan(0-10): ____
42.	Tuşların üstünde bilmediğim dilde yazılar varsa	Puan(0-10): ____
43.	Alette bilmediğim bir dil kullanılıyorsa	Puan(0-10): ____
44.	Kullanılan dil açık değilse	Puan(0-10): ____
45.	Satın alırken açıklayıcı bilgi verilmezse	Puan(0-10): ____
46.	Satıcı nasıl kullanacağımı göstermezse	Puan(0-10): ____
47.	Bilen kişilere sorma şansım yoksa	Puan(0-10): ____
48.	Bilen biri tarafından kullanım anlatılmazsa	Puan(0-10): ____
49.	Kullanımı gösterecek biri yoksa	Puan(0-10): ____
50.	Zorlandığımda yardım alabileceğim biri yoksa	Puan(0-10): ____
51.	Kullanabilen birini gözlemleme şansım yoksa	Puan(0-10): ____
52.	Yardım alabileceğim kimse yoksa	Puan(0-10): ____
53.	Takıldığım zaman yardım edecek kimse yoksa	Puan(0-10): ____
54.	Kılavuzu yoksa	Puan(0-10): ____
55.	Kılavuzda verilen bilgiler net değilse	Puan(0-10): ____

3/5

Puanlama		
Aleti öğrenebileceğime kesinlikle **güvenmiyorum**	0 1 2 3 4 5 6 7 8 9 10	Aleti öğrenebileceğime kesinlikle **güveniyorum**

56.	Kılavuz yeterince açıklayıcı değilse	Puan(0-10): ____
57.	Kılavuz anlaşılamıyorsa	Puan(0-10): ____
58.	Kullanım kılavuzu yeterince anlaşılır değilse	Puan(0-10): ____
59.	Kullanım kılavuzu açıklayıcı değilse	Puan(0-10): ____
60.	Kullanım kılavuzunda yalın bir dil yoksa	Puan(0-10): ____
61.	Kullanım kılavuzunda günlük dilde kullanılmayan sözcükler bulunuyorsa	Puan(0-10): ____
62.	Kılavuzda teknik terimler kullanılıyorsa	Puan(0-10): ____
63.	Teknik servisten telefonla yardım almak mümkün değilse	Puan(0-10): ____
64.	İstediğim kadar deneme yapma şansım yoksa	Puan(0-10): ____
65.	Kullanabilmek için önce sayfalarca kılavuz okumam gerekiyorsa	Puan(0-10): ____
66.	Bir kaç kez kullandığımda hala sorun yaşıyorsam	Puan(0-10): ____
67.	İlk kullanımda sorun yaşarsam	Puan(0-10): ____
68.	Kullanırken çok hata yapıyorsam	Puan(0-10): ____
69.	Aleti sıkça kullanma fırsatı bulamıyorsam	Puan(0-10): ____
70.	Yanımda zaten o aleti kullanmayı üstlenmiş biri varsa	Puan(0-10): ____
71.	Denerken aletin bozulma ihtimali varsa	Puan(0-10): ____
72.	Yanlış yaptığımda geri dönüş yoksa	Puan(0-10): ____
73.	Çabuk arızalanacak bir alet olduğunu düşünüyorsam	Puan(0-10): ____
74.	Alette kullanılan kısaltmaların ne anlama geldiğini bilmiyorsam	Puan(0-10): ____
75.	Daha önceden alet hakkında bilgim yoksa	Puan(0-10): ____
76.	Zor kontrol edilen bir aletse	Puan(0-10): ____
77.	Kullanmadan önce bir sürü ayar yapmak gerekiyorsa	Puan(0-10): ____
78.	İlk kez açıldığında ayarlanması gereken çok şey varsa	Puan(0-10): ____
79.	Yaptıklarımın doğru mu yanlış mı olduğunu anlamakta zorlanıyorsam	Puan(0-10): ____
80.	Hangi tuşa basınca ne olduğu açık değilse	Puan(0-10): ____
81.	Kullanımı mantığıma uygun değilse	Puan(0-10): ____
82.	Alet yaptıklarımı iptal etme şansı vermiyorsa	Puan(0-10): ____
83.	Ciddi sonuçlara yol açabilecek hata yapma ihtimali varsa	Puan(0-10): ____
84.	Kullanım sırasında bir çok şeyi aklımda tutmam gerekiyorsa	Puan(0-10): ____

4/5

Puanlama		
	0 1 2 3 4 5 6 7 8 9 10	
Aleti öğrenebileceğime kesinlikle **güvenmiyorum**		Aleti öğrenebileceğime kesinlikle **güveniyorum**

85.	Kullanım sırasında gerekli bilgileri alet bana hatırlatmıyorsa	Puan(0-10): ____
86.	Ekranda önemli bilgiler net olarak verilmiyorsa	Puan(0-10): ____
87.	Menülerde ihtiyacımdan çok daha fazla bilgi veriliyorsa	Puan(0-10): ____
88.	Hata uyarıları anlaşılmıyorsa	Puan(0-10): ____
89.	Hata uyarıları beni çözüme yönlendirmiyorsa	Puan(0-10): ____
90.	Hata oluştuğunda nedeni anlaşılamıyorsa	Puan(0-10): ____
91.	Ekranda bir sürü gereksiz bilgi varsa	Puan(0-10): ____
92.	Alet karışık ekranlara sahipse	Puan(0-10): ____

Lütfen kullandığınız elektronik aletleri işaretleyiniz.

Cep telefonu	O	Dijital fotoğraf makinası	O	Video kamera	O	MP3 çalar	O
Elektronik ajanda	O	Elektronik sözlük	O	DVD/VCD çalar	O	DVD/VCD kaydedici	O
Müzik seti	O	Televizyon	O	Elektronik oyun	O	Uydu alıcı	O
Bankamatik	O						

Lütfen çevrenizdeki insanlardan (ailenizdekilerden, arkadaşlarınızdan, akrabalarınızdan) **daha iyi** kullandığınızı düşündüğünüz elektronik aletleri işaretleyiniz.

Cep telefonu	O	Dijital fotoğraf makinası	O	Video kamera	O	MP3 çalar	O
Elektronik ajanda	O	Elektronik sözlük	O	DVD/VCD çalar	O	DVD/VCD kaydedici	O
Müzik seti	O	Televizyon	O	Elektronik oyun	O	Uydu alıcı	O
Bankamatik	O						

Çalışmaya katıldığınız için çok teşekkür ederiz…

5/5

APPENDIX K

ITEM-REMAINDER COEFFICIENTS AFTER MAJOR DATA COLLECTION

Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Item no.	26	59	58	60	53	52	48	24	90	61	56	55	50	6	86
Item-remainder coefficient	0,86	0,85	0,85	0,85	0,84	0,83	0,83	0,83	0,83	0,83	0,83	0,83	0,83	0,83	0,83
Rank	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Item no.	27	25	23	41	4	49	30	28	3	76	51	54	57	85	92
Item-remainder coefficient	0,82	0,82	0,82	0,82	0,82	0,82	0,82	0,82	0,82	0,82	0,82	0,81	0,81	0,81	0,81
Rank	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
Item no.	22	47	80	34	44	5	33	29	73	91	89	79	88	45	46
Item-remainder coefficient	0,81	0,81	0,81	0,81	0,81	0,81	0,80	0,80	0,80	0,80	0,80	0,80	0,80	0,80	0,80
Rank	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Item no.	9	66	7	63	74	37	75	82	42	67	38	78	81	10	77
Item-remainder coefficient	0,80	0,79	0,79	0,79	0,79	0,79	0,79	0,79	0,79	0,78	0,78	0,78	0,78	0,78	0,78
Rank	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
Item no.	62	84	68	39	43	32	8	69	71	83	72	64	40	2	1
Item-remainder coefficient	0,78	0,78	0,78	0,77	0,77	0,77	0,77	0,76	0,76	0,76	0,76	0,76	0,76	0,75	0,75
Rank	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
Item no.	21	13	31	18	12	11	35	65	87	19	36	14	15	17	16
Item-remainder coefficient	0,75	0,74	0,74	0,73	0,73	0,72	0,71	0,71	0,71	0,71	0,71	0,69	0,68	0,68	0,67
Rank	91	92*													
Item no.	20	70													
Item-remainder coefficient	0,67	0,45													

*items with an item-remainder coefficient below 0.65

APPENDIX L

FACTOR LOADINGS AFTER PRINCIPAL COMPONENT ANALYSIS

	Components								
ITEMS	1	2	3	4	5	6	7	8	9
1	0,31	0,68	0,18	0,18	0,25	0,16	0,23	0,04	-0,01
2	0,25	0,73	0,22	0,16	0,27	0,18	0,13	0,08	0,07
3	0,30	0,71	0,20	0,22	0,28	0,19	0,16	0,10	0,12
4	0,24	0,67	0,27	0,28	0,24	0,23	0,16	0,15	0,06
5	0,24	0,70	0,21	0,24	0,23	0,23	0,17	0,17	0,12
6	0,26	0,69	0,30	0,23	0,21	0,26	0,17	0,09	0,11
7	0,26	0,72	0,22	0,25	0,18	0,28	0,17	0,04	0,08
8	0,25	0,68	0,24	0,17	0,16	0,23	0,28	0,06	0,14
9	0,23	0,65	0,17	0,31	0,22	0,24	0,16	0,19	0,18
10	0,28	0,59	0,17	0,31	0,18	0,18	0,26	0,11	0,19
11	0,30	0,34	0,16	0,51	0,18	0,07	0,23	0,03	0,40
12	0,22	0,30	0,22	0,52	0,22	0,08	0,20	0,04	0,49
13	0,20	0,31	0,21	0,54	0,25	0,09	0,23	0,05	0,41
14	0,20	0,28	0,18	0,51	0,22	0,04	0,22	0,08	0,47
15	0,20	0,25	0,18	0,67	0,21	0,32	0,05	-0,02	0,09
16	0,23	0,18	0,15	0,74	0,23	0,19	0,11	0,01	0,09
17	0,21	0,22	0,17	0,74	0,24	0,20	0,14	0,07	-0,10
18	0,16	0,17	0,30	0,72	0,17	0,17	0,23	0,17	0,06
19	0,19	0,14	0,26	0,75	0,15	0,13	0,26	0,10	0,12
20	0,21	0,19	0,25	0,69	0,09	0,17	0,16	0,14	0,02
21	0,15	0,28	0,22	0,63	0,18	0,12	0,37	0,19	0,08
22*	0,14	0,34	0,32	0,43	0,22	0,25	0,38	0,21	0,14
23*	0,17	0,37	0,29	0,33	0,22	0,28	0,49	0,21	0,13
24*	0,21	0,40	0,25	0,31	0,26	0,36	0,44	0,13	0,09
25	0,16	0,37	0,31	0,29	0,24	0,30	0,51	0,20	0,13
26*	0,17	0,41	0,30	0,31	0,29	0,33	0,43	0,15	0,18
27*	0,21	0,33	0,38	0,24	0,23	0,35	0,45	0,10	0,16
28	0,28	0,24	0,35	0,29	0,24	0,19	0,54	0,19	0,13
29	0,26	0,27	0,25	0,26	0,30	0,21	0,62	0,15	0,06
30	0,27	0,26	0,25	0,29	0,30	0,22	0,60	0,17	0,03
31	0,23	0,22	0,19	0,27	0,29	0,21	0,54	0,28	-0,04

32	0,35	0,22	0,29	0,32	0,16	0,15	0,56	0,08	0,14
33	0,36	0,29	0,29	0,25	0,24	0,24	0,54	0,01	0,08
34*	0,34	0,23	0,36	0,28	0,17	0,30	0,44	0,07	0,23
35	0,18	0,32	0,23	0,19	0,19	0,69	0,15	0,12	0,03
36	0,17	0,26	0,15	0,20	0,26	0,71	0,18	0,14	0,00
37	0,28	0,27	0,20	0,23	0,30	0,63	0,27	0,05	0,06
38	0,29	0,27	0,19	0,23	0,32	0,62	0,23	0,05	0,06
39	0,32	0,29	0,17	0,21	0,28	0,55	0,28	0,07	0,10
40	0,22	0,41	0,16	0,20	0,27	0,56	0,26	0,14	-0,03
41*	0,35	0,25	0,35	0,27	0,27	0,49	0,25	-0,03	0,16
42*	0,34	0,13	0,37	0,28	0,24	0,29	0,48	-0,03	0,18
43*	0,29	0,11	0,48	0,23	0,27	0,37	0,38	-0,13	0,20
44*	0,32	0,16	0,47	0,23	0,30	0,38	0,37	-0,12	0,16
45	0,24	0,22	0,29	0,22	0,58	0,20	0,35	0,09	0,16
46	0,21	0,24	0,28	0,21	0,70	0,20	0,25	0,13	0,10
47	0,21	0,30	0,24	0,23	0,67	0,28	0,19	0,13	0,14
48	0,23	0,27	0,28	0,30	0,70	0,22	0,17	0,14	0,08
49	0,25	0,27	0,26	0,27	0,74	0,22	0,15	0,12	0,09
50	0,25	0,25	0,29	0,26	0,71	0,21	0,18	0,13	0,12
51	0,25	0,28	0,33	0,19	0,67	0,29	0,19	0,10	0,03
52	0,26	0,29	0,34	0,20	0,65	0,26	0,19	0,13	0,07
53	0,25	0,32	0,29	0,20	0,64	0,23	0,28	0,13	0,10
54	0,24	0,25	0,71	0,29	0,26	0,15	0,22	0,07	0,14
55	0,24	0,28	0,72	0,28	0,22	0,21	0,19	0,10	0,07
56	0,26	0,28	0,72	0,19	0,27	0,19	0,25	0,12	0,06
57	0,27	0,24	0,72	0,26	0,27	0,16	0,19	0,12	0,06
58	0,30	0,28	0,69	0,29	0,29	0,14	0,18	0,15	0,04
59	0,29	0,25	0,68	0,29	0,30	0,16	0,20	0,13	0,10
60	0,31	0,25	0,62	0,26	0,31	0,12	0,30	0,21	0,03
61	0,32	0,30	0,53	0,29	0,29	0,16	0,22	0,21	0,01
62*	0,30	0,27	0,48	0,19	0,28	0,21	0,27	0,16	0,06
63	0,28	0,24	0,56	0,19	0,32	0,25	0,19	0,15	0,09
64	0,30	0,17	0,53	0,22	0,17	0,31	0,14	0,24	0,24
65*	0,21	0,29	0,37	0,13	0,27	0,36	0,15	0,37	-0,06
66*	0,33	0,26	0,46	0,22	0,24	0,21	0,27	0,35	-0,02
67*	0,33	0,31	0,36	0,23	0,34	0,15	0,19	0,44	-0,04
68*	0,38	0,35	0,27	0,18	0,32	0,10	0,25	0,49	-0,02
69*	0,37	0,25	0,38	0,15	0,28	0,13	0,24	0,46	0,07
71*	0,34	0,23	0,37	0,15	0,30	0,20	0,18	0,47	0,19
72*	0,40	0,13	0,40	0,22	0,23	0,14	0,18	0,46	0,32

73*	0,44	0,19	0,41	0,26	0,20	0,18	0,18	0,42	0,25
74	0,55	0,29	0,25	0,21	0,24	0,28	0,11	0,19	0,26
75*	0,49	0,33	0,25	0,18	0,27	0,26	0,19	0,17	0,22
76	0,53	0,29	0,23	0,25	0,19	0,31	0,21	0,18	0,31
77*	0,45	0,35	0,31	0,21	0,15	0,44	0,08	0,17	0,12
78*	0,44	0,40	0,27	0,20	0,17	0,44	0,09	0,15	0,11
79	0,53	0,36	0,30	0,16	0,12	0,36	0,17	0,14	0,22
80	0,59	0,29	0,35	0,15	0,18	0,33	0,15	0,12	0,18
81	0,59	0,25	0,30	0,27	0,21	0,20	0,10	0,13	0,27
82	0,52	0,16	0,36	0,30	0,21	0,15	0,19	0,23	0,32
83	0,54	0,16	0,32	0,26	0,15	0,18	0,21	0,22	0,33
84	0,54	0,39	0,25	0,11	0,20	0,24	0,20	0,27	0,03
85	0,60	0,39	0,19	0,22	0,25	0,18	0,28	0,15	-0,01
86	0,64	0,34	0,20	0,31	0,27	0,18	0,17	0,14	0,03
87	0,54	0,46	0,23	0,16	0,19	0,19	0,09	0,14	-0,17
88	0,68	0,28	0,29	0,23	0,18	0,16	0,30	0,02	0,01
89	0,70	0,26	0,26	0,22	0,22	0,13	0,31	0,05	0,04
90	0,64	0,20	0,34	0,26	0,25	0,16	0,32	0,05	0,11
91	0,54	0,36	0,24	0,30	0,34	0,22	0,15	0,06	-0,12
92	0,54	0,38	0,21	0,28	0,35	0,27	0,15	0,08	-0,09

Extraction Method: Principal Component Analysis.
*Items that do not significantly (above 0.50) load any components

APPENDIX M

FACTORS AND CORRESPONDING ITEMS

Factor 1 – Good interface design

74	Alette kullanılan kısaltmaların ne anlama geldiğini bilmiyorsam
76	Zor kontrol edilen bir aletse
79	Yaptıklarımın doğru mu yanlış mı olduğunu anlamakta zorlanıyorsam
80	Hangi tuşa basınca ne olduğu açık değilse
81	Kullanımı mantığıma uygun değilse
82	Alet yaptıklarımı iptal etme şansı vermiyorsa
83	Ciddi sonuçlara yol açabilecek hata yapma ihtimali varsa
84	Kullanım sırasında bir çok şeyi aklımda tutmam gerekiyorsa
85	Kullanım sırasında gerekli bilgileri alet bana hatırlatmıyorsa
86	Ekranda önemli bilgiler net olarak verilmiyorsa
87	Menülerde ihtiyacımdan çok daha fazla bilgi veriliyorsa
88	Hata uyarıları anlaşılmıyorsa
89	Hata uyarıları beni çözüme yönlendirmiyorsa
90	Hata oluştuğunda nedeni anlaşılamıyorsa
91	Ekranda bir sürü gereksiz bilgi varsa
92	Alet karışık ekranlara sahipse

Factor 2 - Familiarity

1	Daha önce aynı işe yarayan bir aleti kullanmadıysam
2	Daha önce karşılaşmadığım bir aletse

3	Daha önceden kullandığım aletlere benzemiyorsa
4	Önceki aletlerden kazandığım tecrübeyi kullanamıyorsam
5	Daha önce kullandığım aletlerden çok farklıysa
6	Diğer aletlerden alıştığım kullanım şeklini uygulayamıyorsam
7	Daha önce alıştığım aletlerle arasında çok fark varsa
8	Kullanımı yaygın olmayan bir aletse
9	Daha önceki modelleriyle benzerlik taşımıyorsa
10	Kullanmaya alışık olmadığım teknolojiler içeriyorsa

Factor 3 – Instruction manual - support

54	Kılavuzu yoksa
55	Kılavuzda verilen bilgiler net değilse
56	Kılavuz yeterince açıklayıcı değilse
57	Kılavuz anlaşılamıyorsa
58	Kullanım kılavuzu yeterince anlaşılır değilse
59	Kullanım kılavuzu açıklayıcı değilse
60	Kullanım kılavuzunda yalın bir dil yoksa
61	Kullanım kılavuzunda günlük dilde kullanılmayan sözcükler bulunuyorsa
63	Teknik servisten telefonla yardım almak mümkün değilse
64	İstediğim kadar deneme yapma şansım yoksa

Factor 4 – Affection - usefulness

11	İlgi alanıma girmiyorsa

12	Bana ilgi çekici gelmediyse
13	Severek aldığım bir alet değilse
14	Kullanmaktan sıkılıyorsam
15	Kullanmayacağım özellikleri varsa
16	İşime yaramayacak özellikleri çoksa
17	Tüm özelliklerini kullanmayacaksam
18	Fazla ihtiyaç duymadığım bir aletse
19	İşime yarayacak bir alet değilse
20	Yaptığım işleri daha iyi yapmamı sağlamayacaksa
21	Sıkça kullanacağım bir alet değilse

Factor 5 – Help from others

45	Satın alırken açıklayıcı bilgi verilmezse
46	Satıcı nasıl kullanacağımı göstermezse
47	Bilen kişilere sorma şansım yoksa
48	Bilen biri tarafından kullanım anlatılmazsa
49	Kullanımı gösterecek biri yoksa
50	Zorlandığımda yardım alabileceğim biri yoksa
51	Kullanabilen birini gözlemleme şansım yoksa
52	Yardım alabileceğim kimse yoksa
53	Takıldığım zaman yardım edecek kimse yoksa

Factor 6 - Complexity

35	Tuşlar birden fazla işe yarıyorsa
36	Çok fazla tuşu varsa
37	Menüsü çok karışıksa
38	Çok karmaşık özelliklere sahipse
39	Alet karmaşıksa
40	Çok fazla özelliğe sahipse

Factor 7 – Intutiveness

25	Çok kullanılan özelliklerini bulmak kolay değilse
28	Hızlı bir şekilde istediğime ulaşamıyorsam
29	Tuşların üstünde ne işe yaradıkları yazmıyorsa
30	Tuşların üstündeki resimler belirgin değilse
31	Sık sık kılavuza başvurmam gerekiyorsa
32	Mantık yürüterek çözebileceğim bir alet değilse
33	Temel özelliklerin nasıl kullanılacağı açık değilse
42	Tuşların üstünde bilmediğim dilde yazılar varsa (.483)

Items with loadings below .50

Nasıl kullanılacağı açık değilse
Kullanımı zor geliyorsa
Aletin kullanımı karışıksa
Kullanımı akılda kalıcı değilse

Çalışma biçimini kavrayamadıysam
Kendi kendime çözmem mümkün değilse
Kullanılan teknik kelimeler anlaşılmıyorsa
Tuşların üstünde bilmediğim dilde yazılar varsa
Alette bilmediğim bir dil kullanılıyorsa
Kullanılan dil açık değilse
Kılavuzda teknik terimler kullanılıyorsa
Kullanabilmek için önce sayfalarca kılavuz okumam gerekiyorsa
Bir kaç kez kullandığımda hala sorun yaşıyorsam
İlk kullanımda sorun yaşarsam
Kullanırken çok hata yapıyorsam
Aleti sıkça kullanma fırsatı bulamıyorsam
Yanımda zaten o aleti kullanmayı üstlenmiş biri varsa
Denerken aletin bozulma ihtimali varsa
Yanlış yaptığımda geri dönüş yoksa
Çabuk arızalanacak bir alet olduğunu düşünüyorsam
Daha önceden alet hakkında bilgim yoksa
Kullanmadan önce bir sürü ayar yapmak gerekiyorsa
İlk kez açıldığında ayarlanması gereken çok şey varsa

APPENDIX N

GISE-S (Final Form)

Elektronik Aletlerin Öğrenimi Anketi

Cinsiyet: Kadın **O** Erkek **O**

Yaş:

Katılacağınız bu çalışmayla, Orta Doğu Teknik Üniversitesi, Endüstri Ürünleri Tasarımı Bölümü'nde yürütülen bir doktora tezine bilimsel katkıda bulunacaksınız. Çalışmanın hiçbir ticari firma ve kuruluşla ilgisi yoktur. Sonuçlar sadece bilimsel amaçla değerlendirilecektir. Çalışmanın gerçekleştirilmesi için gerekli tüm izinler resmi mercilerden alınmıştır. Çalışmaya katılanların kimlik bilgileri tamamen gizli tutulacaktır. Çalışma bireylerin elektronik aletleri öğrenme konusundaki yaklaşımlarını ölçmek için bir anket geliştirilmesini amaçlamaktadır. Bu ankette **elektronik alet** veya **alet** olarak tanımlanan ürünler:

Cep telefonları;
Dijital fotoğraf makinaları, video kameralar;
Dijital ekranlı fırın, çamaşır makinası, bulaşık makinası gibi ürünler;
MP3 çalarlar;
Elektronik ajandalar, sözlükler;
DVD/VCD oynatıcılar ve kaydediciler, uydu alıcılar;
Müzik setleri;
Televizyonlar;
Elektronik oyunlar;
Bankamatik cihazları...

...gibi **ekran ve tuşlar aracılığıyla** kullanılan ürünlerdir. Lütfen soruları bu tanıma giren ürünleri düşünerek cevaplayınız.

Anket formunda, sorularda elektronik alet öğrenimi sırasında karşılaşılabilecek bazı durumlar anlatılmaktadır. Sorularda anlatılan durumları düşününüz. Anlatılan durumu yaşasaydınız bir elektronik aleti öğrenme konusunda kendinize ne kadar güvenirdiniz? Değerlendirmeyi bir elektronik aleti öğrenebilme konusunda kendinize duyduğunuz güvene göre yapınız. Bu amaçla aşağıda verilen puanlamayı kullanınız.

Soruları 0-10 arasında puan vererek yanıtlayınız. Kendinize güvenmiyorsanız düşük, güveniyorsanız yüksek puanlar veriniz

	0	1	2	3	4	5	6	7	8	9	10
Aleti öğrenebileceğime kesinlikle **güvenmiyorum**											Aleti öğrenebileceğime kesinlikle **güveniyorum**

Soruların nasıl cevaplanacağına ilişkin bir örnek aşağıda verilmiştir:

Örnek: Aşağıdaki durumlarda bir kutuyu kaldırabileceğinize ilişkin güveninizi değerlendiriniz ve 1 ile 10 arasında uygun puanı veriniz.

Kutunun ağırlığı **2** kiloysa	Puan(0-10):	10
Kutunun ağırlığı **5** kiloysa	Puan(0-10):	10
Kutunun ağırlığı **10** kiloysa	Puan(0-10):	10
Kutunun ağırlığı **20** kiloysa	Puan(0-10):	9
Kutunun ağırlığı **40** kiloysa	Puan(0-10):	6
Kutunun ağırlığı **60** kiloysa	Puan(0-10):	3
Kutunun ağırlığı **100** kiloysa	Puan(0-10):	0

1/3

Puanlama

Aleti öğrenebileceğime kesinlikle **güvenmiyorum**	0	1	2	3	4	5	6	7	8	9	10	Aleti öğrenebileceğime kesinlikle **güveniyorum**

1.	Daha önce aynı işe yarayan bir aleti kullanmadıysam	Puan(0-10): ____
2.	Daha önce karşılaşmadığım bir aletse	Puan(0-10): ____
3.	Diğer aletlerden alıştığım kullanım şeklini uygulayamıyorsam	Puan(0-10): ____
4.	Daha önce alıştığım aletlerle arasında çok fark varsa	Puan(0-10): ____
5.	Kullanmaya alışık olmadığım teknolojiler içeriyorsa	Puan(0-10): ____
6.	Severek aldığım bir alet değilse	Puan(0-10): ____
7.	Kullanmaktan sıkılıyorsam	Puan(0-10): ____
8.	İşime yaramayacak özellikleri çoksa	Puan(0-10): ____
9.	Fazla ihtiyaç duymadığım bir aletse	Puan(0-10): ____
10.	Sıkça kullanacağım bir alet değilse	Puan(0-10): ____
11.	Çok kullanılan özelliklerini bulmak kolay değilse	Puan(0-10): ____
12.	Hızlı bir şekilde istediğime ulaşamıyorsam	Puan(0-10): ____
13.	Sık sık kılavuza başvurmam gerekiyorsa	Puan(0-10): ____
14.	Mantık yürüterek çözebileceğim bir alet değilse	Puan(0-10): ____
15.	Temel özelliklerin nasıl kullanılacağı açık değilse	Puan(0-10): ____
16.	Tuşlar birden fazla işe yarıyorsa	Puan(0-10): ____
17.	Çok fazla tuşu varsa	Puan(0-10): ____
18.	Menüsü çok karışıksa	Puan(0-10): ____
19.	Çok karmaşık özelliklere sahipse	Puan(0-10): ____
20.	Alet karmaşıksa	Puan(0-10): ____
21.	Satıcı nasıl kullanacağımı göstermezse	Puan(0-10): ____
22.	Bilen kişilere sorma şansım yoksa	Puan(0-10): ____
23.	Kullanımı gösterecek biri yoksa	Puan(0-10): ____
24.	Kullanabilen birini gözlemleme şansım yoksa	Puan(0-10): ____
25.	Takıldığım zaman yardım edecek kimse yoksa	Puan(0-10): ____
26.	Kılavuzu yoksa	Puan(0-10): ____
27.	Kılavuz yeterince açıklayıcı değilse	Puan(0-10): ____
28.	Kılavuz anlaşılamıyorsa	Puan(0-10): ____

2/3

Puanlama											
Aleti öğrenebileceğime kesinlikle **güvenmiyorum**	0	1	2	3	4	5	6	7	8	9	10 Aleti öğrenebileceğime kesinlikle **güveniyorum**

29.	Kullanım kılavuzunda günlük dilde kullanılmayan sözcükler bulunuyorsa	Puan(0-10): ____
30.	Teknik servisten telefonla yardım almak mümkün değilse	Puan(0-10): ____
31.	Yaptıklarımın doğru mu yanlış mı olduğunu anlamakta zorlanıyorsam	Puan(0-10): ____
32.	Alet yaptıklarımı iptal etme şansı vermiyorsa	Puan(0-10): ____
33.	Ciddi sonuçlara yol açabilecek hata yapma ihtimali varsa	Puan(0-10): ____
34.	Ekranda önemli bilgiler net olarak verilmiyorsa	Puan(0-10): ____
35.	Hata uyarıları anlaşılmıyorsa	Puan(0-10): ____

Lütfen kullandığınız elektronik aletleri işaretleyiniz.

Cep telefonu	**O**	Dijital fotoğraf makinası	**O**	Video kamera	**O**	MP3 çalar	**O**
Elektronik ajanda	**O**	Elektronik sözlük	**O**	DVD/VCD çalar	**O**	DVD/VCD kaydedici	**O**
Müzik seti	**O**	Televizyon	**O**	Elektronik oyun	**O**	Uydu alıcı	**O**
Bankamatik	**O**						

Lütfen çevrenizdeki insanlardan (ailenizdekilerden, arkadaşlarınızdan, akrabalarınızdan) **daha iyi** kullandığınızı düşündüğünüz elektronik aletleri işaretleyiniz.

Cep telefonu	**O**	Dijital fotoğraf makinası	**O**	Video kamera	**O**	MP3 çalar	**O**
Elektronik ajanda	**O**	Elektronik sözlük	**O**	DVD/VCD çalar	**O**	DVD/VCD kaydedici	**O**
Müzik seti	**O**	Televizyon	**O**	Elektronik oyun	**O**	Uydu alıcı	**O**
Bankamatik	**O**						

Çalışmaya katıldığınız için çok teşekkür ederiz…

APPENDIX O

GISE-S (FINAL FORM - ENGLISH)

Learning Electronic Devices Questionnaire

Gender: Female **O** Male **O**

Age:

By participating this study, you will contribute to a PhD thesis conducted in Middle East Technical University, Department of Industrial Design. This study is not done under contract and is not affiliated with any firm. Data gathered will be utilized for scientific purposes. All the official permissons for conducting this study are obtained from legal institutions. Your identity will be kept secret. The study aims to measure the attitude of individuals towards learning new electronic devices. In this questionnaire, the products defined as **electronic devices** or **devices** are products such as:

Cellular phones;
Digital cameras, video cameras;
Products such as digital ovens, washing machines, dishwashers;
Electronic calendars and dictionaries;
DVD/VCD players and recorders, satellite receivers;
HiFi equipment;
Television sets;
Electronic games and consoles;
Automatic Teller Machines (ATMs)...

...that have **buttons and screens.** Please answer the following questions considering this definition.

In the form, circumstances that you may experience while learning a new electronic device are depicted. While answering the questions think about the circumstances defined. If you were to experience that circumstance would you feel confident in learning the electronic device? Please base your evaluation on your confidence in learning a new electronic device. For this purpuse, use the scoring scheme provided below.

Answer the questions by choosing a score between 0-10. If you are not confident score lower, if you feel confident score higher.

	0	1	2	3	4	5	6	7	8	9	10
I definetly don't feel confident that I can learn the device.											**I definetly feel confident** that I can learn the device.

Following example is an exercise before you start filling the actual questionnaire:

Example: Evaluate your confidence in lifting a box under conditions listed below by providing scores between 1-10.

If the box weighs **2** kg	Score(0-10): 10
If the box weighs **5** kg	Score(0-10): 10
If the box weighs **10** kg	Score(0-10): 10
If the box weighs **20** kg	Score(0-10): 9
If the box weighs **40** kg	Score(0-10): 6
If the box weighs **60** kg	Score(0-10): 3
If the box weighs **100** kg	Score(0-10): 0

1/3

Scores

I definetly don't feel confident that I can learn the device.	0	1	2	3	4	5	6	7	8	9	10	**I definetly feel confident** that I can learn the device.

1.	If it is not a type of device that I uses before	Score(0-10):____
2.	If it is a type of device that I didn't use before	Score(0-10):____
3.	If I can't apply the style of use that I learnt using other devices	Score(0-10):____
4.	If it has many differences with a device that I used to	Score(0-10):____
5.	If it has technologies that I am not familiar with	Score(0-10):____
6.	If it is not a product that I liked and bought	Score(0-10):____
7.	If I get bored of using the device	Score(0-10):____
8.	If it has many features that I do not need	Score(0-10):____
9.	If I don't need the product much	Score(0-10):____
10.	If it is not a device that I frequently use	Score(0-10):____
11.	If it is not easy to find the most frequently used functions	Score(0-10):____
12.	If I cannot quickly access what I want	Score(0-10):____
13.	If I often have to refer to instruction manual	Score(0-10):____
14.	If it is not a device that I can work out simply by reasoning	Score(0-10):____
15.	If basic functions are not easy to use	Score(0-10):____
16.	If buttons have more than one function	Score(0-10):____
17.	If it has many buttons	Score(0-10):____
18.	If it has a complex menu	Score(0-10):____
19.	If it has complicated features	Score(0-10):____
20.	If device is complex	Score(0-10):____
21.	If seller does not show me how to use it	Score(0-10):____
22.	If I do not have the opportunity to ask people who know the product	Score(0-10):____
23.	If there is no one to show how to it	Score(0-10):____
24.	If I do not have the opportunity to observe someone while using the product	Score(0-10):____
25.	If there is nobody to help me when I got stuck	Score(0-10):____
26.	If it does not have an instruction manual	Score(0-10):____
27.	If manual is not sufficiently descriptive	Score(0-10):____
28.	If manual is hard to comprehend	Score(0-10):____

2/3

Scores		
I definetly don't feel confident that I can learn the device.	0 1 2 3 4 5 6 7 8 9 10	**I definetly feel confident** that I can learn the device.

29.	If there are words in the manual that are not used in everyday language	Score(0-10):____
30.	If I cannot get assistance from call center	Score(0-10):____
31.	If I can hardly understand whether the things I did are right or wrong	Score(0-10):____
32.	If device does not give me the opportunity to cancel what I do	Score(0-10):____
33.	If there is the possibility to make a mistake that may cause serious damage	Score(0-10):____
34.	If crucial information is not clearly displayed	Score(0-10):____
35.	If error messages cannot be understood	Score(0-10):____

Please mark the electronic devices you use

Cellular phone	O	Digital camera	O	Video camera	O	MP3 player	O
Electronic calendar	O	Electronic dictionary	O	DVD/VCD player	O	DVD/VCD recorder	O
HiFi Equipment	O	Television	O	Elektronic games	O	Satellite receiver	O
ATM	O						

Please mark the devices which you think that **you use better** than people around you (family members, friends, relatives)

Cellular phone	O	Digital camera	O	Video camera	O	MP3 player	O
Electronic calendar	O	Electronic dictionary	O	DVD/VCD player	O	DVD/VCD recorder	O
HiFi Equipment	O	Television	O	Elektronic games	O	Satellite receiver	O
ATM	O						

Thanks...

3/3

APPENDIX P

GISE-S LITE AFTER SEM

Elektronik Aletlerin Öğrenimi Anketi

Cinsiyet: Kadın **O** Erkek **O**

Yaş:

Çalışmaya katılanların kimlik bilgileri tamamen gizli tutulacaktır. Çalışma bireylerin elektronik aletleri öğrenme konusundaki yaklaşımlarını ölçmek için bir anket geliştirilmesini amaçlamaktadır. Bu ankette **elektronik alet** veya **alet** olarak tanımlanan ürünler:

Cep telefonları;
Dijital fotoğraf makinaları, video kameralar;
Dijital ekranlı fırın, çamaşır makinası, bulaşık makinası gibi ürünler;
MP3 çalarlar;
Elektronik ajandalar, sözlükler;
DVD/VCD oynatıcılar ve kaydediciler, uydu alıcılar;
Müzik setleri;
Televizyonlar;
Elektronik oyunlar;
Bankamatik cihazları...

...gibi **ekran ve tuşlar aracılığıyla** kullanılan ürünlerdir. Lütfen soruları bu tanıma giren ürünleri düşünerek cevaplayınız.

Anket formunda, sorularda elektronik alet öğrenimi sırasında karşılaşılabilecek bazı durumlar anlatılmaktadır. Sorularda anlatılan durumları düşününüz. Anlatılan durumu yaşasaydınız bir elektronik aleti öğrenme konusunda kendinize ne kadar güvenirdiniz? Değerlendirmeyi bir elektronik aleti öğrenebilme konusunda kendinize duyduğunuz güvene göre yapınız. Bu amaçla aşağıda verilen puanlamayı kullanınız.

Soruları 0-10 arasında puan vererek yanıtlayınız. Kendinize güvenmiyorsanız düşük, güveniyorsanız yüksek puanlar veriniz

	0	1	2	3	4	5	6	7	8	9	10
Aleti öğrenebileceğime kesinlikle **güvenmiyorum**											Aleti öğrenebileceğime kesinlikle **güveniyorum**

Soruların nasıl cevaplanacağına ilişkin bir örnek aşağıda verilmiştir:

Örnek: Aşağıdaki durumlarda bir kutuyu kaldırabileceğinize ilişkin güveninizi değerlendiriniz ve 1 ile 10 arasında uygun puanı veriniz.

Kutunun ağırlığı **2** kiloysa	Puan(0-10): ________
Kutunun ağırlığı **5** kiloysa	Puan(0-10): ________
Kutunun ağırlığı **10** kiloysa	Puan(0-10): ________
Kutunun ağırlığı **20** kiloysa	Puan(0-10): ________
Kutunun ağırlığı **40** kiloysa	Puan(0-10): ________
Kutunun ağırlığı **60** kiloysa	Puan(0-10): ________

1/2

Puanlama		
	0 1 2 3 4 5 6 7 8 9 10	
Aleti öğrenebileceğime kesinlikle **güvenmiyorum**		Aleti öğrenebileceğime kesinlikle **güveniyorum**

Kutunun ağırlığı **100** kiloysa Puan(0-10): ________

1.	Daha önce aynı işe yarayan bir aleti kullanmadıysam	Puan(0-10): ____
2.	Daha önce karşılaşmadığım bir aletse	Puan(0-10): ____
3.	Diğer aletlerden alıştığım kullanım şeklini uygulayamıyorsam	Puan(0-10): ____
4.	Severek aldığım bir alet değilse	Puan(0-10): ____
5.	Kullanmaktan sıkılıyorsam	Puan(0-10): ____
6.	Fazla ihtiyaç duymadığım bir aletse	Puan(0-10): ____
7.	Çok kullanılan özelliklerini bulmak kolay değilse	Puan(0-10): ____
8.	Mantık yürüterek çözebileceğim bir alet değilse	Puan(0-10): ____
9.	Temel özelliklerin nasıl kullanılacağı açık değilse	Puan(0-10): ____
10.	Tuşlar birden fazla işe yarıyorsa	Puan(0-10): ____
11.	Menüsü çok karışıksa	Puan(0-10): ____
12.	Alet karmaşıksa	Puan(0-10): ____
13.	Satıcı nasıl kullanacağımı göstermezse	Puan(0-10): ____
14.	Bilen kişilere sorma şansım yoksa	Puan(0-10): ____
15.	Kullanımı gösterecek biri yoksa	Puan(0-10): ____
16.	Kılavuz anlaşılamıyorsa	Puan(0-10): ____
17.	Kullanım kılavuzunda günlük dilde kullanılmayan sözcükler bulunuyorsa	Puan(0-10): ____
18.	Teknik servisten telefonla yardım almak mümkün değilse	Puan(0-10): ____
19.	Yaptıklarımın doğru mu yanlış mı olduğunu anlamakta zorlanıyorsam	Puan(0-10): ____
20.	Alet yaptıklarımı iptal etme şansı vermiyorsa	Puan(0-10): ____
21.	Ciddi sonuçlara yol açabilecek hata yapma ihtimali varsa	Puan(0-10): ____

Çalışmaya katıldığınız için çok teşekkür ederiz…

2/2

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name:	Berkman, Ali Emre
Nationality:	Turkish (TC)
Date and Place of Birth:	December 15, 1976, Ankara
Marital Status:	Married
Phone:	+90 312 444 62 66
Fax:	+90 312 210 18 72
Email:	ali.berkman@gmail.com

EDUCATION

Degree	Institution	Year of Graduation
MS	METU Industrial Design	2002
BS	METU Industrial Design	1998
High School	Kolej Ayşeabla	1994

WORK EXPERIENCE

Year	Place	Enrollment
2008 - Present	UTRLAB User Testing and Research	Director of User Research
2002 - 2008	METU/BiltirUTEST	Usability Expert
1999 - 2006	METU Department of Industrial Design	Research Assistant
1996 - 1997	METU Department of Industrial Design	Student Assistantship
1996 July	Altı Tasarım	Intern Design Student
1995 July	Aselsan	Intern Design Student

FOREIGN LANGUAGES

Advanced English

PUBLICATIONS

1. Tamer, A., Karapars, Z. Akar, E., Berkman A.E., Sel Kaygın, S. (2010). "User research for the challenges of convergence on designing next generatıon TVs". In: NMIC 2010 - 2nd International Conference on New Media and Interactivity, April 28-30, Istanbul, Turkey.

2. Berkman, A.E. (2009) General Interaction Expertise and General Interaction Self-Efficacy: A Multi-view Approach to Sampling in Usability Testing of Consumer Products, *Human Computer Interaction* (Ioannis Pavlidis Editor), IN-Tech: Vienna.

3. Vermeeren, A.P.O.S., Attema, J., Akar, E., Ridder, H., Van Doorn, A. K., Erbuğ, Ç., Berkman, A. E., Maguire, M. (2008). Usability Problem Reports for Comparative Studies: Consistency and Inspectability, *Human Computer Interaction,* 23 (4), pp. 329-380.

4. Berkman, A. E. (2003). Existing and potential accessibility of private bathroom spaces in Turkey. *Proceedings of the international conference: CIB W062 2003 water drainage and supply systems.*

5. Berkman, A. E. & Erbuğ, Ç. (2005). Accommodating individual differences in usability studies on consumer products. *Proceedings of the 11th conference on human computer interaction,* Volume 3.

6. Erbuğ, Ç., Vermeeren, A.P.O.S., Berkman, A. E., Akar, E., McDonagh, D. (2005). Usability testing: a collaborative approach. *Proceedings of the 11th conference on human computer interaction,* Volume 3.

7. Berkman, A.E., (2007). General Interaction Expertise: An Approach for Sampling in Usability Testing of Consumer ProductsJ. Jacko (Ed.): Human Computer Interaction, Volume I, HCII 2007 pp. 397-406, Springer: Berlin.