

EVALUATION OF STUDENT INFORMATION SYSTEM (SIS) IN TERMS OF USER
EMOTIONS, PERFORMANCE, AND PERCEIVED USABILITY: A TURKISH
UNIVERSITY CASE

DENİZHAN DEMİRKOL

JUNE, 2018

EVALUATION OF STUDENT INFORMATION SYSTEMS (SIS) IN TERMS OF
USER EMOTIONS, PERFORMANCE AND PERCEIVED USABILITY:

A TURKISH UNIVERSITY CASE

BY

DENİZHAN DEMİRKOL

Supervised By:

Assistant Professor Dr. Çağla ŞENELER

DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ARTS
IN
DEPARTMENT OF MANAGEMENT INFORMATION SYSTEMS

YEDİTEPE UNIVERSITY

JUNE, 2018

APPROVAL PAGE

Approval of the Institute of Social Sciences



Professor Dr. M. Fazıl Güler

Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Arts.



Professor Dr. Avadis Simon Hacınlıyan

Head of Department

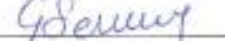
This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Arts.



Assistant Professor Dr. Çağla Şeneler

Supervisor

Examining Committee Members

Assistant Professor Dr. Çağla Şeneler – Yeditepe University 

Assistant Professor Dr. Serdar Çiftçi – Adnan Menderes University 

Associate Professor Dr. Aşkın Demirağ – Yeditepe University 

PLAGIRASIM PAGE**PLAGIARISM**

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.

Date : 07.06.2018

Name/Surname: Denizhan DEMİRKOL

Signature :



ABSTRACT

Nowadays especially students are increasingly demanding online access to their databases, and universities, colleges are working to meet these demands. Universities and various organizations provide their services via information systems. Student Information Systems (SIS) are fundamental part of academic systems in higher education. SIS are case in point trying to provide information for registered students in courses, documenting grading, transcripts, results of students tests and other assessment scores, building student schedules, tracking student attendance and managing many other student-related data needs in a school. SIS are used by many diverse types of student groups. This programme of research investigates whether different student groups' background as a group difference can affect system users' emotion, performance, and perceived usability by conducting a pilot and main study.

To investigate user emotion, the Emotion Word Prompt List (EWPL) was used. The EWPL was developed in English, making it unsuitable for Turkish learners, the target student groups for this research programme. Therefore, the EWPL was translated into Turkish by requesting a permission from the main author of the EWPL who is Professor Dr. Helen Petrie via professional techniques and adequate reliability and validity established by administering it 324 undergraduate students in Turkey. Henceforth, the Turkish version of the EWPL will be referred as the EWPL-TR.

In order to measure perceived usability, the System Usability Scale (SUS) was used. The SUS was developed in English, making it unsuitable for Turkish learners, the target student groups for this research programme. Therefore, the SUS was translated into Turkish by requesting a permission from main author of the SUS who is Dr. John Brooke, through professional techniques and adequate reliability and validity established by administering it 324 undergraduate students in Turkey. Henceforth, the Turkish version of the SUS will be referred as the SUS-TR.

While developing EWPL-TR and SUS-TR, demographic data is also collected from participants in order to evaluate different student groups' emotion and perceived usability about the SIS. The results of this study showed that participants' emotional reactions to SIS, and likes and dislikes of the aspects of SIS differ with their specific demographic backgrounds.

The aim of pilot study was before experimenting large student groups to identify possible problems. Pilot test was conducted to six participants. It has been found interesting results. To illustrate, female participants showed more negative emotions to SIS compared to male participants.

The final study investigated how different student groups' backgrounds affect users' emotion, performance and perceptions of the usability of SIS. It was administered in the one of the universities in the Turkey- Istanbul city with an sample of 32 participants. Participants were divided technology prone group and other group to measure the differences of the groups. Participants were asked to perform a number of tasks on the system, the tasking being carefully chosen to direct participants to different aspects of the SIS. In addition, EWPL-TR,

SUS-TR and Overall Assessment Questions were directed to participants. This study showed significant differences among users emotion, perceived usability and performances depending on their specific demographic backgrounds.

The results of this research programme will be important for developers of SIS and similar websites who want to take the advantage of developing supportive systems by focusing on users' demographic backgrounds.

Keywords: Human-computer interaction, student information systems, usability evaluation, system usability scale, emotion word prompt list, translation, reliability, validity



ÖZET

Günümüzde özellikle öğrenciler giderek artan bir şekilde veritabanlarına çevrimiçi erişim talep etmektedir ve üniversiteler ve personelleri bu talepleri karşılamak için çalışmaktadır. Üniversiteler ve çeşitli organizasyonlar bilgi sistemlerini kullanarak hizmet vermektedir. Öğrenci Bilgi Sistemleri (OBS), yüksek öğretimdeki akademik sistemlerin temel parçasıdır. OBS, derslerde kayıtlı öğrencilere bilgi sağlamaya, not verme, transkript, öğrenci testlerinin sonuçlarını ve diğer değerlendirme puanlarını belgelemeye, öğrenci programlarını oluşturmaya, öğrenci katılımını takip etmeye ve bir okuldaki diğer birçok öğrenci veri ihtiyacını yönetmeye yardımcı olmaktadır. OBS çok çeşitli öğrenci grupları tarafından kullanılır. OBS akademik dünyada yaygın olarak çalışılmasına rağmen yapılan çalışmalar bu sistemlerin daha üretken hâle getirilmesi için hala üzerinde çalışılması gerektiğini göstermektedir. Bu boşluğu doldurmak için, bir Türk vakıf üniversitesinin OBS'si, kullanıcı duygusu, performansı ve algılanan kullanılabilirlik açısından değerlendirilmiştir.

Kullanıcı duygularını araştırmak için, Emotion Word Prompt List (EWPL) kullanılmıştır. EWPL, İngilizce olarak geliştirilmiştir ve Türk öğrenciler bu araştırma programının hedef öğrenci grupları için uygun değildir. Bu nedenle EWPL, EWPL'nin ana yazarı Profesör Dr. Helen Petrie'den izin istemek suretiyle, profesyonel tekniklerle, Türkiye'de 324 lisans öğrencisi tarafından oluşturulan yeterli güvenilirlik ve geçerlilik çalışmaları ile Türkçe'ye çevrilmiştir. Bundan sonra, EWPL'nin Türkçe versiyonu EWPL-TR olarak anılacaktır.

Algılanan kullanılabilirliği ölçmek için Sistem Kullanılabilirlik Ölçeği (SUS) kullanılmıştır. SUS, İngilizce olarak geliştirilmiştir ve Türk öğrenciler bu araştırma programının hedef öğrenci grupları için uygun değildir. Bu nedenle SUS, SUS'un ana yazarı Dr. John Brooke tarafından izin istemek suretiyle, profesyonel tekniklerle, Türkiye'de 324 lisans öğrencisi tarafından oluşturulan yeterli güvenilirlik ve geçerlilik çalışmaları ile Türkçe'ye çevrilmiştir. Bundan sonra, SUS'un Türkçe versiyonu SUS-TR olarak anılacaktır.

EWPL-TR ve SUS-TR'yi geliştirirken, farklı öğrenci gruplarının duygularını ve OBS hakkındaki algılanan kullanılabilirliği değerlendirmek için katılımcılardan demografik veriler de toplanmıştır. Bu çalışmanın sonuçları, katılımcıların OBS'ye karşı duygusal tepkilerinin ve algılanan kullanılabilirliğinin belirli demografik geçmişlerine göre farklılaştığını göstermiştir.

Pilot çalışmanın amacı büyük öğrenci gruplarını denemeden önce, olası problemleri tanımlamaktır. Pilot test altı katılımcıya yapılmıştır ve ilginç sonuçlar bulunmuştur. Örneğin, kadın katılımcılar erkek katılımcılara kıyasla OBS'ye daha olumsuz duygular göstermiştir.

Son çalışma, farklı öğrenci gruplarının demografik özelliklerinin, kullanıcıların duygularını, performansını ve OBS'nin kullanılabilirlik algılarını nasıl etkilediğini araştırmıştır. Türkiye'de İstanbul ilindeki üniversitelerinden birinde 32 katılımcıdan oluşan bir örneklem ile çalışma gerçekleştirilmiştir. Katılımcılar gruplar arasındaki farklılıkları tespit etmek için teknoloji eğilimli grup ve diğer grup olarak iki gruba ayrılmıştır. Katılımcılardan sistemde bir dizi görev yapmaları istenmiş ve görevler, OBS'nin farklı yönlerine yönlendirilmek için dikkatli bir şekilde seçilmiştir. Ayrıca katılımcılara EWPL-TR, SUS-TR ve Genel Değerlendirme Anketi yöneltilmiştir. Bu çalışma, spesifik demografik özelliklere bağlı olarak, kullanıcıların duyguları, algılanan kullanılabilirlik ve performansları arasında anlamlı farklılıklar olduğunu göstermiştir.

Bu araştırma programının sonuçları, kullanıcıların demografik geçmişlerine odaklanarak destekleyici web siteleri geliştirmekten yararlanmak isteyen OBS ve benzeri sistemlerin geliştiricileri için önemli olacaktır.

Anahtar kelimeler: İnsan-bilgisayar etkileşimi, kullanılabilirlik değerlendirmesi, öğrenci bilgi sistemleri, sistem kullanılabilirlik ölçeği, duygusal kelime istem listesi, çeviri, geçerlilik, güvenilirlik



To my supportive family...



ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my advisor Dr. Çağla Şeneler for the continuous support of my master study and related research, for her patience, motivation, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis. I owe gratitude to my counselor, whom I learned a lot in a scientific and academic sense. I could not have imagined having a better advisor and mentor for my study. I hope that our relationship based on sympathy and respect between us will continue to grow stronger throughout life. Additionally, I want to thank Güçlü Şeneler who support my research.

My sincere thanks also go to Professor Dr. Avadis Simon Hacınlıyan and Associate Professor Dr. Askın Demirağ, who provided me an opportunity to conduct a research, and who gave access to the laboratory and research facilities. Without their precious support, it would not be possible to conduct this research. Also, I would like to thank Dr. Halise Şerefoğlu for her support and facilitate my thesis studies. I would like to thank Dr. Asil Alkaya for his valuable recommendations. Furthermore, I want to express my sincere thanks to Professor Dr. M. Fazıl Güler for his support. Besides, I want to represent my special thanks to Professor Dr. Mustafa Çetin and Professor Dr. Fatma Neval Genç for their tolerance during my research.

I would also like to thank Dr. Serdar Çiftci who encourage my studies. Separately I would like to express my intimate gratitude to Professor Dr. Yücel Bozdağlıoğlu for enlightening and illuminating me the first glance of research.

Last but not least, I am deeply thankful to my family. Special thanks go to my mother Nurten Demirkol and my father Mehmet Sıraç Demirkol for their love, patience, motivation and encouragement. I dedicate this work to them.

TABLE OF CONTENTS

APPROVAL PAGE.....	i
PLAGIRASIM PAGE.....	ii
ABSTRACT	iii
ÖZET.....	v
DEDICATION.....	vii
ACKNOWLEDGEMENTS.....	viii
TABLE OF CONTENTS.....	ix
LIST OF TABLES	xiv
LIST OF FIGURES	xviii
1. INTRODUCTION.....	1
1.1. Background	1
1.2. Research Issues	4
2. LITERATURE REVIEW	6
2.1. Introduction to Literature Review	6
2.2. Definition and Types of Information Systems (IS)	6
2.2.1. Student Information Systems (SIS).....	7
2.3. Human-Computer Interaction (HCI) Concepts	8
2.3.1. Usability	9
2.3.2. Usability Measurement- Usability Evaluation Methods	10
2.3.3. Usability Evaluation Phases	15
2.3.3. User Experience and User Emotion	19

2.3.4. User Performance	21
2.4. Examples of Studies About Usability of the SIS in the Literature.....	21
2.5. Examples of Studies About User Experience and Emotion of the Information Systems in the Literature	22
2.6. Discussion and Conclusions	23
3. TRANSLATION OF THE EMOTION WORD PROMPT LIST INTO TURKISH, THE EWPL-TR.....	24
3.1. Introduction	24
3.2. Methodology.....	25
3.2.1. Translation method of the EWPL into Turkish.....	25
3.2.2. Reliability and validity study of the Turkish version of the EWPL (EWPL-TR).....	27
3.3. Findings	39
3.4. Discussion	40
4. TRANSLATION OF THE SUS INTO TURKISH, THE SUS-TR.....	42
4.1. Introduction	42
4.1.1. The Distinction between the SUS-TR and Kadirhan, Gül and Battal's Study	45
4.2. Translation Of The SUS Into Turkish.....	46
4.2.1. Translation Method Of The SUS Into Turkish.....	46
4.2.2. Reliability and Validity Study of the SUS-TR.....	48
4.3. Findings	58
4.5. Discussion	60
5. INVESTIGATING THE EFFECTS OF DIFFERENT STUDENT GROUPS ON USERS' EMOTION AND PERCEIVED USABILITY OF SIS	61
5.1. Method.....	61
5.2. Participants	61
5.4. Questionnaires	67
5.4.1. EWPL-TR.....	67

5.4.2. SUS-TR	67
5.4.3. Overall Reactions Questionnaire (ORQ)	67
5.4.4. Demographic Questionnaire	67
5.4.5. Procedure	67
5.5. Data Analysis.....	68
5.5.1. Student Background	68
5.5.2. Emotion Word Prompt List (EWPL-TR) Data.....	69
5.5.3. System Usability Scale (SUS-TR) Data	69
5.5.4. Overall Reactions Questionnaire (ORQ) Answers	70
5.6. Effects of Different Student Groups on Users' Emotion.....	71
5.7. Effects of Different Student Groups on Users' Perceived Usability	90
5.8. Relation Between EWPL-TR and SUS-TR.....	102
5.9. Discussion and Conclusion.....	103
6. PILOT STUDY.....	108
6.1. Introduction	108
6.2. Study Context.....	108
6.2.1. Organizational Setting.....	108
6.2.2. Objective	108
6.3. Method.....	109
6.3.1. Study Design	109
6.3.2. Participants.....	109
6.3.3. Test Scenarios.....	110
6.3.4. Usability Analysis.....	110
6.3.5. Outcome Measures.....	111
6.4. Results.....	112
6.5. Discussion	113
6.5.1. Limitations of the study.....	113

6.6. Conclusion	113
7. EVALUATION OF STUDENT INFORMATION SYSTEM (SIS) IN TERMS OF USER EMOTION, PERFORMANCE AND PERCEIVED USABILITY.....	115
7.1. Introduction	115
7.2. Method	115
7.2.1. Design	115
7.2.2. Participants.....	118
7.2.3. Equipment	123
7.2.4. SIS and Tasks	123
7.2.5. Questionnaires.....	125
7.2.6. Procedure.....	126
7.3. Data Analysis.....	126
7.3.1. Student Background	127
7.3.2. Emotion Word Prompt List (EWPL-TR) Data.....	127
7.3.3. System Usability Scale (SUS-TR) Data	127
7.3.4. Overall Reactions Questionnaire (ORQ) Answers	127
7.3.5. Effects of Different Student Groups on Users’ Emotion, Performance and Perceived Usability on SIS	127
7.4. Discussion and Conclusion.....	141
8. DISCUSSION AND CONCLUSIONS	145
8.1. Synthesis of Main Results from this Programme of Research	145
8.2. Limitations of the studies	146
8.3. Future work.....	147
8.4. Implications for SIS developers.....	148
REFERENCES	149
APPENDIX A: List of Abbreviation	156
APPENDIX B.....	157
APPENDIX C	161

C.1. Öğrenci Bilgi Sistemi Görevler Listesi.....	161
C.2. Gönüllü Katılım Formu.....	162
C.3. Öğrenci Bilgi Sistemi Genel Değerlendirme Anketi	163
C.4. Kullanıcı Bilgi Anketi.....	164
APPENDIX D	166
D.1 GLOSSARY	166
APPENDIX E	167
DENİZHAN DEMİRKOL	167



LIST OF TABLES

Table 3.1. : <i>Demographic Information of the Working Group</i>	28
Table 3.2. : <i>Web Usage, Computer Usage and Confidence Statistics of the Participants</i>	29
Table 3.3. : <i>Item-total Statistics of the EWPL-TR</i>	30
Table 3.4. : <i>Item-total statistics (after component extraction)</i>	31
Table 3.5. : <i>Reliability Coefficients</i>	32
Table 3.6. : <i>KMO and Bartlett test results</i>	32
Table 3.7.: <i>Total variance explained</i>	33
Table 3.8. : <i>Varimax-rotated two-factor result for the EWPL-TR</i>	35
Table 3.9. : <i>Positive EWPL-TR Factor</i>	36
Table 3.10. : <i>Negative EWPL-TR Factor</i>	36
Table 3.11. : <i>EWPL-TR model results</i>	38
Table 3.12. : <i>Values of Compliance Criteria for the EWPL Scale Cfa Model</i>	39
Table 4.1. : <i>Demographic Information of the Working Group</i>	49
Table 4.2. : <i>Web Usage, Computer Usage and Confidence Statistics of the Participants</i>	50
Table 4.3. : <i>Item-total Statistics of the SUS-TR</i>	51
Table 4.4. : <i>Reliability coefficients</i>	51
Table 4.5. : <i>KMO and Bartlett's Test Results</i>	52
Table 4.6. : <i>Total Variance Explained</i>	53
Table 4.7.: <i>Varimax-rotated Two-factor Solution for the Turkish Version of the System Usability Scale</i>	55
Table 4.8. : <i>Positive SUS-TR Factor</i>	55
Table 4.9. : <i>Negative SUS-TR Factor</i>	56
Table 4.10. : <i>SUS-TR Scale Model Results</i>	57
Table 4.11. : <i>Values of compliance criteria for the SUS-TR scale CFA model</i>	58
Table 5.1. : <i>Demographic Information of the Working Group</i>	62
Table 5.2. : <i>Education Demographics of Participants</i>	63
Table 5.3. : <i>Participants' Computeral Demographic Characteristics</i>	64
Table 5.4. : <i>Participants' Categorized Overall Reaction Questionnaire Responses</i> ...	66
Table 5.5. : <i>Varimax-rotated Two-Factor Result for the EWPL-TR</i>	69
Table 5.6. : <i>Category Set for Overall Reactions</i>	70
Table 5.7. : <i>Normal Distribution Test Results of the Participants' EWPL-TR Scale</i> ..	71
Table 5.8. : <i>Descriptive Statistics of Participants' EWPL-TR Scale</i>	71
Table 5.9. : <i>Independent Sampling T-Test Results of Differentiation of EWPL Subscales According to Gender</i>	72
Table 5.10. : <i>Results of One-Way ANOVA Test for Differentiation of EWPL Subscales According to Age Groups</i>	72
Table 5.11. : <i>One-Way ANOVA Test Results for EWPL-TR Scale measured Under Different Duration Sizes at the University</i>	74
Table 5.12. : <i>Independent Sampling T-Test Results for Status of the EWPL-TR Scale Under different Registered Departments</i>	75
Table 5.13. : <i>One-way ANOVA test results for differences of EWPL-TR Sub-Scales according to the overall GPA</i>	76
Table 5.14. : <i>One-Way ANOVA Test Results for Differentiation of Dimensions of EWPL-TR Scale according to Computer Usage Time</i>	78

Table 5.15. : <i>One-Way ANOVA Test Results for Differentiation of the EWPL-TR Sub-Scales with respect to the Computer Usage Skills</i>	79
Table 5.16. : <i>One-way ANOVA Test Results for the Differentiation of Dimensions of EWPL-TR Scale According to Self-Confidence Variation on Computer Usage</i>	80
Table 5.17. : <i>Independent Sampling T-Test Results For The Differentiation of EWPL-TR Scale Based On The Internet Web Service Usage Frequency</i>	81
Table 5.18. : <i>One-way ANOVA test results for differentiation of dimensions of the EWPL-TR scale based on the University' SIS Usage Frequency</i>	83
Table 5.19. : <i>One-way ANOVA Test Results for Differentiation of EWPL-TR Sub-Scales according to the Most Liked Features of SIS at the University</i>	85
Table 5.20. : <i>One-way ANOVA Test Results for Differentiation of EWPL-TR Subscales under the Most Disliked Features of SIS at the University</i>	87
Table 5.21. : <i>One-way ANOVA Test Results of Differentiation of the EWPL-TR Subscales according to Recommendations about Increase Usage of SIS at the University</i>	89
Table 5.22. : <i>Normal Distribution Test Results of the Participants' SUS-TR Scale</i>	90
Table 5.23. : <i>Descriptive Statistics of Participants' SUS-TR Scale</i>	91
Table 5.24. : <i>Independent Sampling T-Test Results for Differentiation Status of the SUS Scale According to the Gender</i>	91
Table 5.25. : <i>One-way ANOVA test results for differentiation of SUS-TR scale according to age groups</i>	92
Table 5.26. : <i>One-Way ANOVA Test Results for SUS-TR Scale measured Under Different Duration Sizes at the University</i>	92
Table 5.27. : <i>Independent Sampling T-Test Results for the SUS-TR Scale Under different Registered Departments</i>	93
Table 5.28. : <i>One-way ANOVA test results for differences of SUS-TR Scale according to the overall GPA</i>	94
Table 5.29. : <i>One-Way ANOVA Test Results for SUS-TR Scale according to Computer Usage Time</i>	94
Table 5.30. : <i>One-Way ANOVA Test Results for Differentiation of the SUS-TR Scale with respect to the Computer Usage Skills</i>	95
Table 5.31. : <i>One-way ANOVA Test Results for the SUS-TR Scale According to Self-Confidence Variation on Computer Usage</i>	96
Table 5.32. : <i>Independent Sampling T-Test Results For The Differentiation of SUS-TR Scale Based On The Internet Web Service Usage Frequency</i>	96
Table 5.33. : <i>One-way ANOVA test results for differentiation of the SUS-TR scale based on the University' SIS Usage Frequency</i>	97
Table 5.34. : <i>One-way ANOVA Test Results for Differentiation of SUS-TR Scale according to the Most Liked Features of SIS at the University</i>	98
Table 5.35. : <i>One-way ANOVA Test Results for Differentiation of the SUS-TR Scale according to the Most Disliked Features of SIS at the University</i>	99
Table 5.36. : <i>One-way ANOVA Test Results of Differentiation of the SUS-TR Scale according to Recommendations about Increase Usage of SIS at the University</i>	100
Table 5.37. : <i>Determination of the Difference Between Age and SUS Score Groups</i>	101
Table 5.38. : <i>Determination of the Differences Between Participants' SUS-TR Score Groups and Most Disliked Features of SIS at the University</i>	101
Table 5.39. : <i>Participants' Categorized SUS-TR Scores</i>	102

Table 5.40. : <i>Pearson Correlation Analysis for the Dependence between the subscales of EWPL and SUS-TR measurements</i>	103
Table 7.1. : <i>Demographic Information of the Working Group</i>	118
Table 7.2. : <i>Participants' Educational Demographics</i>	119
Table 7.3. : <i>Computer Usage Characteristics of the Participants in the Experimental Study</i>	120
Table 7.4. : <i>Participants' Categorized Overall Reactions Questionnaire Responses</i>	122
Table 7.5. : <i>Artificial Tasks and Test Scenarios</i>	124
Table 7.6. : <i>Descriptive Statistics of the Participants by EWPL-TR and SUS subscales Based on Questionnaires</i>	128
Table 7.7. : <i>Mann Whitney U Test Results of EWPL-TR Subscales according to Different Student Groups</i>	128
Table 7.8. : <i>Kruskal Wallis H Test results for Differentiation of EWPL-TR Subscales based on the Most Liked Feature Group of SIS</i>	129
Table 7.9. : <i>Kruskal Wallis H Test Results for SUS-TR Scales according to the Most Liked Feature Group in the SIS</i>	131
Table 7.10. : <i>Kruskal Wallis H Test Results for SUS Subscales According to Class Groups</i>	132
Table 7.11. : <i>Mann Whitney U Test Results of Task Success in the System and Differences between IT and Other Group</i>	133
Table 7.12. : <i>Mann Whitney U Test Results of Net Times Spent on Tasks according to System Differences between IT and Other Group</i>	135
Table 7.13. : <i>Mann Whitney U Test Results of Number of Clicks for the Tasks according to System Differences between IT and Other Group</i>	136
Table 7.14. : <i>Relationship Between the Time Spent and Number of Clicks by IT and Other Group</i>	136
Table 7.15. : <i>Pearson Correlation Analysis Results between the different dimension of the EWPL-TR perceptions of the participants and Mean Net Time, Mean number of clicks and mean number of successful tasks</i>	137
Table 7.16. : <i>Pearson Correlation Analysis Results between SUS-TR scores of the participants and Mean Net Time, Mean number of clicks and mean number of successful tasks</i>	138
Table 7.17. : <i>Independent Sample T-Test Results of Mean Net Time, Mean number of Clicks and Mean number of Successful Tasks according to their Gender</i>	139
Table 7.18. : <i>Independent Sample T-Test Results of Participants' Mean Net Time, Mean number of clicks and Mean number of Successful Tasks according to Self-Confidence variation on computer usage</i>	140
Appendix Table B.1. : <i>Original version of the EWPL-</i>	157
Appendix Table B.2. : <i>Finalized Turkish Version of the EWPL (EWPL-TR)</i>	158
Appendix Table B.3. : <i>Original Version of the SUS</i>	159
Appendix Table B.4. : <i>Turkish Version of the SUS (SUS-TR)</i>	160

Appendix Table C.1. : <i>Turkish version of the tasks according to SIS</i>	161
--	-----



LIST OF FIGURES

<i>Figure 2.1. : Functions of an Information System (IS)</i>	7
<i>Figure 2.2. : Definition of usability by Shackel (1991)</i>	9
<i>Figure 2.3. : Ten Heuristics for System Design (Nielsen,1995)</i>	13
<i>Figure 2.4. : “In a heat-map, a screenshot is colour-coded according to the number of looks each part attracts”</i>	14
<i>Figure 3.1. : First Step of the Interpretation Method</i>	26
<i>Figure 3.2. : Second Step of the Interpretation Method</i>	27
<i>Figure 3.3. : The Third Step of the Interpretation Method</i>	27
<i>Figure 3.4. : Scree Plot Test Graph</i>	34
<i>Figure 3.5. : CFA Results for EWPL-TR</i>	37
<i>Figure 4.1. : The first step of the Interpretation Process</i>	47
<i>Figure 4.2. : Second Step of the Interpretation Process</i>	47
<i>Figure 4.3. : Third Step of the Interpretation Process</i>	48
<i>Figure 4.4. : Scree plot test graph</i>	54
<i>Figure 4.5. : CFA results for SUS-TR scale</i>	57
<i>Figure 7.1. :A sample of the IT Group Population</i>	116
<i>Figure 7.2. : A sample of the Other Group Population</i>	117
<i>Figure 7.3. : Task Success Results for Information Technology Group Students</i>	133
<i>Figure 7.4. : Task Success Results for Other Group Students</i>	134

1. INTRODUCTION

1.1. Background

Internet technology has now become an integral part of our lives. People use web sites for the purpose of getting information, social networking, communication, entertainment and online education. Nowadays especially students are increasingly demanding online access to their databases, and universities, colleges are working to meet these demands.

Universities and various organizations provide their services via websites. Student Information Systems (SIS) are fundamental part of academic systems in higher education. SIS is a valid example attempting to give data to enrolled students in courses, archiving grading, transcripts, results of students tests and other appraisal scores, building student plans, following student participation and checking numerous other student related information needs in a school.

Many persons such as currently studying and graduated students, academics, faculty secretaries and officers use these services. For this reason, usability is extremely important in the design of SIS (Şengel, 2014).

Diverse classes of students now utilize SIS for numerous sources of information. Notwithstanding, the diversity and complexity of the online information available on SIS and the desire to have systems provide all information for all users, regardless their interest, ability or characteristics, means that information systems can be devastating to users. Effective SIS not only satisfies personals and students but also increase the sustainability of academic workflow. In order to prevent user errors, maintain users' tasks and increase users' satisfaction towards to the system, usability study is essential for SIS. Usually, SIS have been developing without usability analysis and user-centered design principles although these efforts should be carried out by developers. Formal usability testing of the SIS could enhance

the framework and increment its valuable consequences for scholastic work process. Although people use educational institutions services, little research and empirical study exist that examine the impact of users's emotion, performance and perceived usability of Student Information Systems (SIS). This study will help to fill this gap. Therefore one foundation University' SIS which is in Turkey – Istanbul city have been evaluated in terms of students emotion, performance and perceived usability.

This study investigates SIS users' emotion, experience, performance and perceived usability by handling two surveys and usability study of users' experience with SIS.

In order to examine users' emotional responses, The Emotion Word Prompting List (EWPL) was selected since it has been previously used in the literature evaluating users' emotions. It has been wished to include Turkish students in this study programme. However, the EWPL was developed in English. Therefore, the EWPL was interpreted through professional techniques into Turkish and the studies on reliability and validity of the survey were performed.

To investigate users' perceived usability of the SIS, System Usability Scale (SUS) was selected because it has been widely used in evaluating usability. However, the SUS was developed in English. Therefore, the SUS was translated via professional techniques into Turkish and the studies on reliability and validity of the survey were carried out. In the course of conducting this work, it was discovered that the SUS had already been interpreted into Turkish. However, it has been found various dilemmas with that interpretation. Unlike the previous interpretation, it has been used professional interpretation techniques during the development of the survey. Also, it has been established the reliability and validity of the survey by receiving permission was obtained from the main author of the SUS who is Dr John

Brooke. Therefore this research programme contributes beneficial as well as a useful Turkish version of the SUS.

After translating the scales to be used in Turkish with professional techniques and completing the validity and reliability studies, a pilot study was conducted to see possible problems. Within the scope of the student information systems study, a pretest voluntary participation form was given to the participants in order to record and analyze demographic information. Consequently, the usability tests were implemented with two groups of real system users. A screen recording instrument was used during the usability studies so as to record and control the screen and calculate the click numbers of the members. The members were chosen among selected University's first cycle students. The first group is prone to technology departments which are Management Information Systems and Information Systems and Technologies students. The second group is rest of the Faculty of Commerce' students. Each member undertook a series of tasks on selected University' SIS. These tasks were carefully constructed to guide participants to elements of the SIS that would students have difficulty in carrying out. After completing the tasks on each website, they finished a series of short surveys (a) their urgent emotional response to the system (using the Emotional Word Prompt List, EWPL); (b) their perception of the usability of the system (as assessed by the SUS, Brooke, 1996); (c) their like/dislike on the systems (as assessed by the Overall Reactions Questionnaire, ORQ). In addition, the click counts and time spent on each task within the selected University' SIS were calculated.

In the light of the above results, SIS study of the research programme investigated in terms of students' emotion, performance and perception of the usability of SIS. Tasks were then generated that would direct people to SIS so that SIS was assessed their performance and experience on SIS in a naturalistic scenario of use. In this way, the emotions, performance and

perceived usability of users with different demographics in selected University's SIS were assessed.

1.2. Research Issues

Research Question: Are there any relations between students' emotion, performance and perceived usability on Student Information Systems?

This research investigates whether any relations between students' emotion, performance and perceived usability of SIS. In order to study this, three research questions have been proposed:

Hypotheses:

H1: The differences of emotions of participants about the SIS according to their characteristics were statistically significant.

H2: The differences of perceived usability of participants about the SIS according to their characteristics were statistically significant.

H3: The differences of task performances of participants about the SIS according to their characteristics were statistically significant.

The remains of the proposal is sorted out as follows:

Chapter 2 introduces related writing and foundation to the thesis. A number of research areas will be covered including SIS, usability definition, usability evaluation methods, user experiences and user emotion and their measurement methods.

Chapter 3 introduces the choosing method for the EWPL used in the research. Furthermore, it provides the development of the Turkish version of the Emotion Word Prompt List (The EWPL-TR) and the works on its reliability and validity.

Chapter 4 introduces the choosing method for the SUS used in the research, Furthermore, it provides the development of the Turkish version of the System Usability Scale (The SUS-TR) and the works on its reliability and validity.

Chapter 5 introduces results of the obtained data from EWPL-TR and SUS-TR which implemented into 324 undergraduate students. In this chapter, it has been examined the effects of different student groups on users' emotion and perceived usability of SIS.

Chapter 6 demonstrates the SIS' pilot study which implemented six university students.

Chapter 7 demonstrates the SIS' main study which investigates that is how demographic background affects users' experience, performance, and perception of the usability of the SIS.

Chapter 8 shows the general discussion of the primary results, commitments, impediments, feasible arrangements and ramifications of the program of research.

2. LITERATURE REVIEW

2.1. Introduction to Literature Review

In the present chapter, literature related to already stated thesis will be covered. Firstly, definitions of information systems (IS) and student information systems (SIS) will be given. Furthermore, up to today examples from academic studies on SIS will be given. After that, human-computer interaction (HCI) definition and related concepts with HCI will be given. Since user experience, emotion and perceived usability are within the scope of this thesis, these concepts will be examined in this chapter in terms of their definitions and evaluation methods. Lastly, this chapter will be concluded with a discussion.

2.2. Definition and Types of Information Systems (IS)

There are various definitions of information systems (IS) in the literature. Some of these definitions are listed in the following:

According to Encyclopedia Britannica; IS can be characterized as a combined arrangement of segments for gathering, depositing, and handling data and for giving information, learning, and advanced items.

Laudon and Laudon (2014) define IS from a technical perspective as it has three fundamental movements that are input, processing, and output which produces the information that institutions require.

There are likewise ecological performing players, for example, clients, providers, contenders, investors and administrative offices, which are converged with the association and its IS. This definition is illustrated in the following figure (See Figure 2.1).

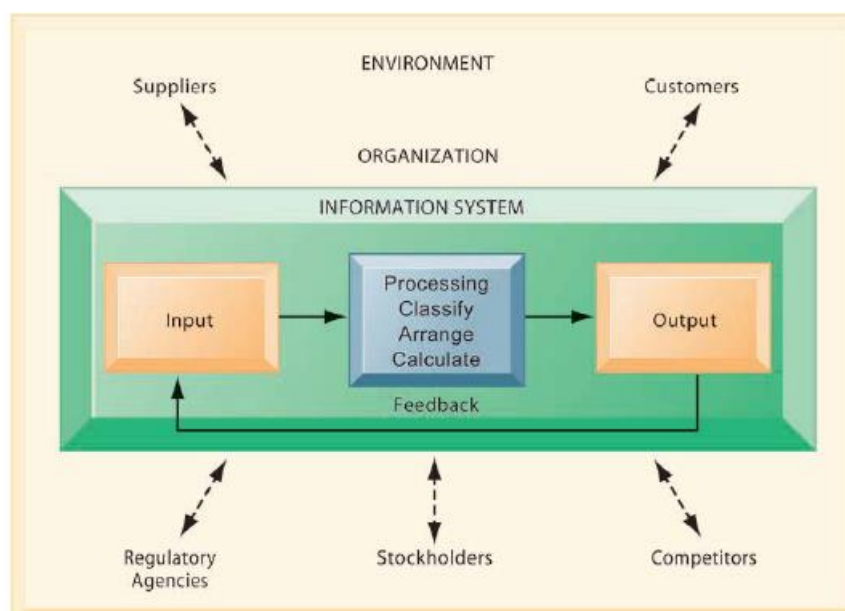


Figure 2.1. : Functions of an Information System (IS) (Laudon & Laudon, 2014 p:47)

Laudon and Laudon (2014) also define IS from a behavioral perspective. Based on this definition IS hold data about noteworthy individuals, spots, and things inside the association or in environment encompassing it (Laudon & Laudon, 2014 p:49).

There are many IS types such as supply chain management systems, customer relationship management systems, data warehouses, enterprise resource management systems, expert systems, geographic information systems, office automation systems, health information systems, learning management systems and student information systems (SIS). SIS is in the scope of this thesis so that will be explained in dept in the following subsection.

2.2.1. Student Information Systems (SIS)

Student information systems (SIS) are a type of IS and it is also known as student management systems, school administration software and student administration systems. SIS is part of typical management information systems.

SIS have lots of functions, such as, giving record administration of previous, current and expected students covering parts of inquiry, application, admission, scholastic execution and past events, permitting charging and expenses management, handling essential occasions,

graduations and examinations, programming classes, announcing operational and administration issues and encouraging information into various other master frameworks (Mukerjee, 2012).

A SIS is essential for any school and integral to its services to students. Today, almost all educational institutions use SIS for recording and covering students' records. SIS is necessary for providing and maintaining student information effectively. It has excessive functions and critical for students, academicians as well as administrative personnel.

2.3. Human-Computer Interaction (HCI) Concepts

The Association for Computing Machinery (ACM) defines "human-computer interaction" (HCI) as "a discipline concerned with the design, evaluation, and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them".

HCI related many disciplines and multidisciplinary subject however, it is in computer-science and frameworks outline that it must be acknowledged as a focal interest (Dix, Finlay, Abowd, & Beale, 2004 p:4). Hornbæk (2006) states that the fundamental research question in HCI is the means by which to work with and enhance the usability of interactive structures.

Research addressing this question has gave rise to guidelines for advancing the usability of systems (Smith & Mosier, 1986), approaches for forecasting usability problems (Molich & Nielsen, 1990), techniques to check the usability of systems (Gould & Lewis, 1985), and discussions on how to measure usability (Nielsen & Levy, 1994; ISO,1998; Frøkjær, Hertzum, & Hornbæk, 2000).

As a result of the core HCI concepts, which are usability, user emotion, user experience and user performances definitions, will be given.

2.3.1. Usability

Usability is progressively perceived as a critical quality factor for interactive software systems, including conventional GUIs- style applications, websites, as well as mobile services. Hornbeak (2006) has defined usability as the main term in HCI field.

ISO standard 9,241-11 Guidance on Usability (1998) defines usability as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use”.

Effectiveness is the “extent to which a goal or task is achieved”, efficiency is the “amount of effort required to accomplish a goal” and satisfaction is the “level of comfort that the users feel when using a product and how acceptable the product is to users as a means of achieving their goals”(Jordan, 1998 p:5).

Usability implies that the general population who utilize the product can do as such rapidly and effectively to finish their own particular tasks. The main consideration of usability is focusing on users (Dumas, S. Jaseph, 1999). If a useful product is desired to be developed, the understanding of potential users is a vital one. Likely the primary generally referred to usability, was set up by and is appeared in the coming figure (See Figure 2.2).

Effectiveness
Improvement in task performance in terms of speed and/or error rate by a given percentage of the population within a given range of the user's tasks (related to the user's environment)
Learnability
Within some specified time from commissioning and start of user training based upon some specified amount of user support and within some specified relearning time each time for intermittent users
Flexibility
With flexibility allowing adaptation to some specified percentage variation in task and/or environments beyond those first specified
Attitude
Within acceptable levels of human cost in terms of tiredness, discomfort, frustration and personal effort, so that satisfaction causes continued and enhanced usage of the system

Figure 2.2. : Definition of usability by Shackel (1991)

According to Jacob Nielsen (1993, 2012) usability is a quality distinguishing that assesses how straightforward user interfaces are to adopt. Besides, the author summarizes five usability attributes that are learnability, efficiency, memorability, errors and user satisfaction. Additionally, Shneiderman and Plaisant (2005) define usability measures which are time to learn, the speed of performance, the rate of errors by users, retention over time and subjective satisfaction.

2.3.2. Usability Measurement- Usability Evaluation Methods

Hornbeak (2006) states that usability cannot be directly measured. For this reason, there are various ways to evaluate usability. These approaches can be classified as expert, theory, and user-based (M.Sweeney, M.Maguire & B.Shackel, 1993).

Expert-based (also called non-empirical) methods include an expert making a judgment about a system's usability or a methodical check of a system's design qualities. However, these methods have considerable disadvantages. To illustrate, it is hard to find trained usability experts. Even if you can find one, these experts are demanding high prices. Furthermore, multiple experts should evaluate the system to aggregate their results (Nielsen, 1992).

Another method to evaluate the product is the model-based approach. The model based approach can be defined as using a model of how a user would use the planned system to acquire supposed system (Sears & Jacko, 2008). However, it has some disadvantages like measuring only some aspects of user performance and inadequate task applicability (Dillon, 2001).

Since systems are developed for users, usability focuses on understanding users. That's why, if developers desire to develop a useful system, the understanding of potential users is vital. Therefore, the last approach that includes user-based (also empirical) methods

focus on observing how users are interacting with systems or allowing users to express their perceptions about systems' usability.

Some of the usability evaluation methods are interviews and questionnaires, thinking aloud method, heuristic evaluation, task analysis, user feedback, eye tracking methodology, summative and formative evaluation.

2.3.2.1. Interviews and Questionnaires

Usability focuses on understanding users. Therefore questionnaires and interviews are appropriate methods for interpreting how users use structures and what characteristics they specifically like or disfavor. These methods are unquestionable related methods because both implicate inquiring users a set of questions and documenting their answers (Nielsen, 1993).

According to Rubin and Chisnel (2014) interviewing is one of the vital elements of usability testing so controlled and large-scale interviewing of the participants by the test moderator is essential. Interviews with individual users can be advantageous since the interviewer can take place of unique topics of burden (Kuhn, 2000). Furthermore, interviews with nonusers of computers would help researchers to figure out the sources of this anxiety and to develop design guidelines for lightening the fear (Shneiderman & Plaisant, 2005).

An assortment of surveys have been utilized and announced in the literature for evaluating the perceived usability of interactive systems, including SUS (System Usability Scale) (Brooke, 1996), QUIS (Questionnaire for User Interface Satisfaction) (Chin, Diehl, & Norman, 1988) and CSUQ (Computer System Usability Questionnaire) (Lewis, 1995). There are also questionnaires, which measure user's emotional reactions to websites. One example of these questionnaires is EWPL (An Emotion Words Priming List) (Petrie & Harrison, 2009).

Questionnaires are generally regulated via mail, and the response rate can be expanded significantly by including a prepaid and pre-addressed reply cover (Nielsen, 1993). Pre-test and post-test questionnaires can be applied to the participant in order to get extensive knowledge about the participant. Pre-test surveys are expected to accumulate product-related data also, not simply demographic information about the member. The primary motivation behind the composed post-test questionnaire(s) is to accumulate inclination data from the members with a specific end goal to clear up and extend your comprehension of the product's qualities and shortcomings (Rubin & Chisnel, 2014).

2.3.2.2. Think Aloud Method

A powerful technique for usability testing is to encourage clients to verbally process about what they are doing. This well-known technique, which is also psychological research method, called "Think – Aloud Method". The planner or analyzer ought to be steady of the members, not assuming control or giving guidelines, but rather provoking and tuning in for intimations about how they are managing the interface. The main advantage of thinking-aloud protocols is gathering clues for the observant usability tester (Shneiderman & Plaisant, 2005).

2.3.2.3. Heuristic Evaluation

Heuristic evaluation is also known as expert evaluation. Generally, these evaluation methods have been carried out by expert usability specialist.

Nielsen and Molich (1990) created the first version of the heuristics, which Nielsen (1995) revised to include the following 10 guidelines, for usability which is given in the following figure (See Figure 2.3).

1. Visibility of system status

The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.

2. Match between system and the real world

The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.

3. User control and freedom

Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.

4. Consistency and standards

Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.

5. Error prevention

Even better than good error messages is a careful design which prevents a problem from occurring in the first place.

6. Recognition rather than recall

Make objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.

7. Flexibility and efficiency of use

Accelerators -- unseen by the novice user -- may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.

8. Aesthetic and minimalist design

Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.

9. Help users recognize, diagnose, and recover from errors

Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.

10. Help and documentation

Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

Figure 2.3. : Ten Heuristics for System Design (Nielsen, 1995).

2.3.2.4. Task Analysis

Task analysis is the way toward finding out about regular users by monitoring them in action to comprehend in detail how they execute their tasks and accomplish their expected aims. Hackos and Redish (1998) note that performing a task analysis helps researchers understand what target users' goals are and what they literally accomplish to conclude their

goals and what experiences users bring to the tasks. Diaper (2004) states task analysis is the center of the nearly all work in HCI because it is related the performance of work and it concentrates on the performance of the system. As a consequence of the way of task analysis, the features of the systems can be evaluated.

2.3.2.5. Eye Tracking Methodology

Eye-tracking method is one of the popular experiments to conduct the usability testing. Equipment generally used in investigations to collect data from a person who takes part in activity about how their bodies react to provocation is being included in comprehensive usability testing (Rubin & Chisnel, 2014). Eye tracking devices enable specialists to see precisely what users detect and where they watch by following their eye activities and the time span they focus on a specific piece of a screen or a word, expression, or object on the screen. (Barnum, 2011).

Heat maps are the most excellent accepted visualization technique for eye tracking experiments (Nielsen & Pernice, 2010) (See Figure 2.4).

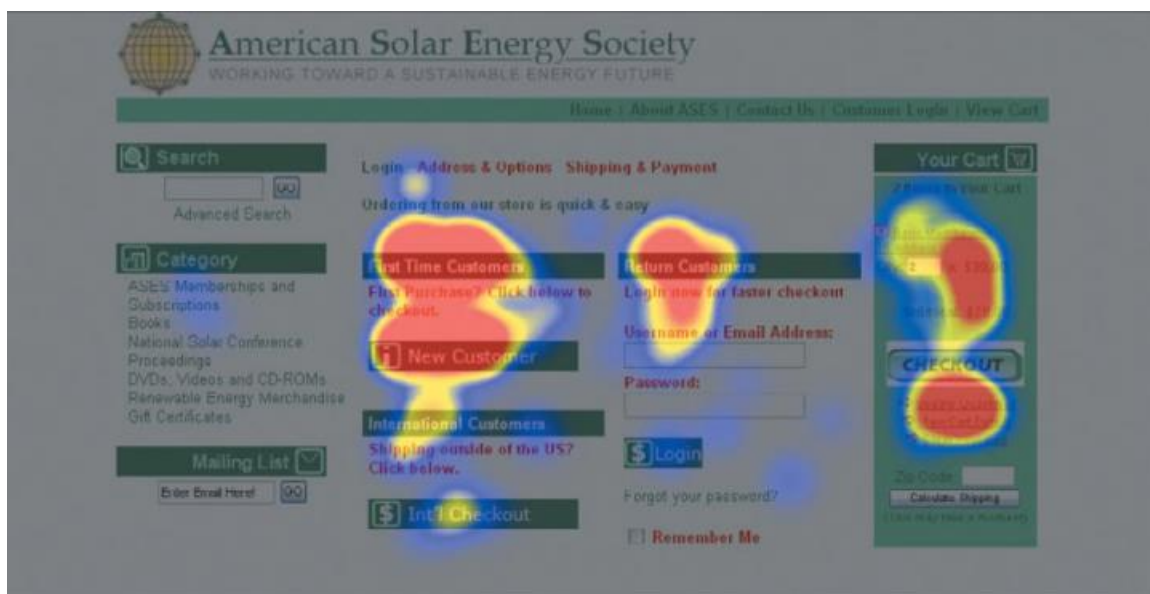


Figure 2.4. : “In a heat-map, a screenshot is colour-coded according to the number of looks each part attracts” (Nielsen & Pernice, 2010)

Eye tracking experiments are extremely trendy with companies focus on where to apportion advertisements on their web pages. Heart rates, blood oxygen levels, and different physiological estimates are fascinating to companies' business accomplices in advertising (Rubin & Chisnel, 2014).

2.3.2.6. Formative and Summative Evaluation

Operated earlier in the development cycle is the formative study-. In thanks to formative evaluation, researchers can acquire user feedback for early designs of software products (Scholtz, 2003). In the light of the data from studies can give rise to the extra user-centred progress of the result (Barnum, 2011).

Usability testing should be done iteratively throughout product development however sometimes researchers must execute the test when the product is almost completed. This kind of testing called summative evaluation since it evaluates the usability of the system or product at the final of the evolution (Barnum, 2011).

2.3.3. Usability Evaluation Phases

Usability evaluation concentrates on how easily users can acquire information and utilize a product in order to conclude their objectives. Furthermore, it covers measurements to understand how satisfied users are with that procedure. Rubin and Chisnell (2008) summarize usability evaluation phases which are developing a test plan, setting up a testing environment, finding and selecting participants, preparing test materials, conducting the test sessions, debriefing the participant and observers, analyzing data and observations and reporting finding and recommendations. These phases will be discussed in the following.

2.3.3.1. Develop a Test Plan

The test design is the establishment of the whole test. Developing a test plan deals with how, when, where, what, and by whom the availability test is conducted. The coming

logics are critical because they are mandatory to expand a detailed test plan. Advancing test plan serves as a draft for the test. The parts of a test design contains reason, objectives, and destinations of the test, research questions, member qualities, strategy (test configuration), assignment list, test condition, hardware, and logistics, test arbitrator part, information to be gathered and assessment measures and report substance and introduction (Rubin & Chisnell, 2008).

2.3.3.2. Set up Testing Environment

Setting up the testing environment is an imperative part of the usability testing. According to Barnum (2011), usability testing can be done in any space such as an office or conference room. There are also usability labs for the tests. A usability lab is a place where users are considered interacting with a system for assessing the system's usability. Although usability laboratories' prices expensive it can be intended to make a perfect testing condition (Barnum, 2011). A common usability lab incorporates a committed room that can be luxuriously fit the client, the arbitrator and if needed maybe a couple others as onlookers, a work area or table and at least two seats for member, for the mediator and for maybe a couple spectators and a PC and Internet access if website or web application was tested. In addition to these a camera to record to the session, a microphone to save everything the participant says additional cameras to get different views from the session(Barnum, 2011). Eye tracking equipment also can be placed to usability labs if wanted.

2.3.3.3. Find and Select Participants

The choice and obtaining of members whose qualifications and capacities are illustrative and presentational of your work's planned user is an indispensable component of the testing series of actions to achieve result. If the general population who accompany are standard clients of the item or as near that standard as could reasonably be expected, your test

outcomes may be constraining and precise. The client profile of the target group of onlookers- portrayal of the behaviors, mastery status, knowledge, demographic info plus additional important variables expected of the general population who will be utilizing the device must be documented. In addition to these, defining the criteria for each user group and specifying prerequisites and classifiers for determination, dividing the target client profile into distinct categories, determining the number of participants to test and scheduling and confirming participants should be carried out. Moreover, participants' privacy and personal information must be protected (Rubin & Chisnell, 2008).

2.3.3.4. Prepare Test Materials

All data collection instruments such as electronic equipment such as cameras, microphones, and computers should be prepared before the participant arrived the test room. Additionally pre-test questionnaire and background information (also known as demographic information), task scenarios, post-task questionnaires should be printed and prepared. Likewise, specific scales such as measure satisfaction or emotions towards to tested system can also be given to the participant.

2.3.3.5. Conduct the Test Sessions

Conducting the test sessions starts from the minute the mitigator welcomes the participant and proceeds to the point that the mitigator closes the meeting, expresses gratitude to the member, and declares farewell. First of all the participant should be met and greeted. After the typical merriments directing the pre-test preparation — utilizing the content so that mitigator say a similar thing each time - can be followed. After that, the participant should be prepared to think aloud. Closure the session by directing your post-test process, which may incorporate a post-test survey, a semi-organized meeting, item response cards or some other subjective input process. Additionally being practical and not prejudiced is highly essential.

Being a perfect mitigator by checking non-verbal communication, balancing adoration, knowing when and how to mediate affects the success of testing (Barnum, 2011).

2.3.3.6. Questioning the Participant and Observers

Questioning alludes to investigating and surveying the user's activities during the action part of a usability test. While the execution of the usability test reveals and uncovered issues, it is frequently the questioning session that reveals insight into why these issues happened and how to settle them. Questioning with observers is would be helpful to look from various perspectives what is going on in the sessions. There are many techniques such as “Devil’s Advocate”, “Replay the Test” and “ What Did You Remember” (Rubin & Chisnell, 2008).

2.3.3.7. Analyze Data and Observations

First of all preliminary analysis should be done to determine the worst problems. These includes compiling, summarizing, analyzing data and developing recommendations. When data have been compiling , the raw data also be organized. Furthermore collected data should be aggregated. It helps analyzing where participants performed well and where they executed unsatisfactorily. Additionally; achievement data in terms of errors and task correctness must be abstracted. Level of members acting effectively, including the individuals who required help and level of members acting effectively within a time benchmark are types of statistics that correlate to task accuracy. Moreover task timings detail to how much time members require to finish each assignment (Rubin & Chisnell, 2008).

Additionally questionnaires that have been collected from participants also should be analyzed.

Identifying the task that did not match the success criterion, recognizing participant errors and dilemmas, conducting a cause of mistaken tasks analysis, prioritizing obstacles, contrasts between gatherings or product forms and utilizing inferential counts are periods of examining process are phases of analyzing process (Rubin & Chisnell, 2008).

The following is an extensive analysis which is more detailed and final report.

2.3.3.8. Report Finding and Recommendations

There are various ways to present your verdicts and Barnum (2011) proposed different options in order to report findings. These are utilizing tables to compress the discoveries, delineating discoveries with screen captures, figures, member statements and video cuts, utilizing visual presentations of qualitative data, for example, the item response cards and making proposals at whatever level of detail is proper and anticipated. In the light of these recommendations increased product 's usability and raised user satisfaction can be expected.

2.3.3. User Experience and User Emotion

HCI is also a field known that worried about understanding, planning for, and assessing a more extensive scope of user experience viewpoints (Sharp, Rogers, & Preece, 2007) and as a result, main motivation who study in this field is to devise interactive systems that are usable and practical (Head, 1999).

The work about of users' emotional behaviour as a vital angle of the user experience has been getting expanding consideration for the past ten years and studying emotions and effect as part of the user experience is essential in order to measure emotional reactions. (Mahlke & Minge, 2008). Palen and Bødker (2008) stated that emotion is an important topic in HCI. User experience has three main typical features which are “user”, “that user is interacting with a product, system, or really anything with an interface” and “the user's experience is of interest, and observable or measurable” (Tullis & Albert, 2013). It is

extremely important to take the informative function of emotions into account when designing HCI interfaces (Kim & Moon, 1998). According to Jordan (1998) creating usable products may not have the same meaning as creating pleasurable products. As a result, it is necessary to measure user emotions. Mahlke and Minge (2008) asserted that this can be performed through various ways such as questionnaires, interviews, various kinds of survey methods or physiological computation such as pupil reactions, heart rate and electrodermal activity, electromyography or the analysis of facial expressions captured by video.

User experience has a huge role in our lives because new products are becoming more and more complex (Tullis & Albert, 2013).

2.3.3.1. How to Measure User Experience?

While some of the evaluation tools mentioned in Section 2.3.2 are already used to evaluate user experience, they were originally designed to evaluate usability. However, it is ambiguous whether they are sufficient to evaluate user experience. Blythe and Monk (2002) offer a positional pyramid as a means of evaluating the user experience of artifacts, using enjoyability, recodification, and inclusivity as measures of user experience.

User experience can be measured by using metrics, which are: task success, user satisfaction, errors, eye tracking data and others. All of the user experience metrics must be “observable” and “quantifiable”. Also user experience metrics are mainly about the communication between the user and product: “some form of effectiveness (being able to complete a task), efficiency (the amount of effort required to complete the task), or satisfaction (the degree to which the user was happy with his or her experience while performing the task)” (Tullis & Albert, 2013 p:7).

2.3.4. User Performance

User performance can be defined “measuring the degree to which users can accomplish a task correctly” (Tullis & Albert, 2013 p:44).

2.3.4.1. How to Measure User Performances?

In order to measure user performance, many measures should be calculated. Some of these are: the time takes to carried out each task, number of mouse clicks and the number of errors acted and how much time a user spends in order to achieve each task (Tullis & Albert, 2013).

2.4. Examples of Studies About Usability of the SIS in the Literature

SIS has become trendy in the last decade and abundant educational institutions have been using SIS. Although there are abundant studies about the usability of universities' library websites (Battleson, Booth, & Weintrop, 2001), there are still few studies about the SIS's usability in the literature.

Among barriers to SIS adoption is poor usability (Carcary, Long & Remenyi, 2007). Usability of the SIS is important because thousands of people including for students, academicians and administrative personnel use SIS in order to conduct educational tasks.

Gemmell and Pagano (2003) studied user perceptions and overall attitude to the effectiveness of a SIS in the Salford University. Data were collected from 84 users through a four-part questionnaire, mostly closed questions. The authors found that the users have experienced difficulties while they are using the system and they indicated that an assessment of the user skill base is required. Furthermore, they stated that users' skill base should be regarded when implementing new functionalities of a system.

The implementation of a new SIS at an Irish Institute of Technology (IoT) is another example. It was found that system usability is identified by nine of the ten individuals interviewed at the point when the issues experienced in the university at the season of framework commissioning. Finally, the Institute has benefited from an advanced system implementation (Carcary et al, 2007).

Another study is carried out in the University Malaysia Sabah (Nordaliela, Suriani, & Nathaniel, 2013). Data was collected using CSUQ. The respondents in that review were 132 students of computer engineering program and two task officers from IT board. Through factor analysis, four factors; beneficial information, timely access, interface design, and error recovery were usability factors which received concerns from the users. Most importantly, they found that usability is a critical factor for user satisfaction.

2.5. Examples of Studies About User Experience and Emotion of the Information Systems in the Literature

There are many conducted studies user experience and emotion of the information systems and websites. For example, Machodo and Tao (2007) investigated the Blackboard LSM and the Moodle LSM. They used online questionnaires to compare the user experience of the key functionalities of the system. They found that the students in the pilot study adopted the Moodle LSM over the Blackboard LSM. The results of the study show that the Moodle LSM was the adopted choice of the students.

Another important study carried out at the University of York. Seneler (2014) investigated learning styles and cultural background on user's experience through comparing two museum websites. It has been discovered that positive feelings of members lead high usability rating of a site and negative feelings lead low usability rating for the site. Furthermore, this research indicated noteworthy contrasts among users relying on their learning styles and cultural backgrounds.

Furthermore, Harrison (2008) investigated emotional web experience by conducting an online survey, builds on the first to explore the role context of use plays in influencing user experience. Responses from 239 people provide means to explore differences in pleasing and frustrating design elements in different web contexts. The results clearly show that the majority of positive feeling words were associated with completion of the task and the negative feelings were being experienced in the earlier stages of the interaction. This shows that websites are as still creating a high level of anxiety to users and this should be tended to lessen the odds of causing work-related pressure.

Despite the fact that the evaluation of many types of information systems in terms of user experience and emotions are measured in the literature, work in SIS is extremely limited.

2.6. Discussion and Conclusions

This chapter introduces the important foundation to help the research considered in this dissertation. Main topics that are discussed in the chapter are SIS, usability, usability evaluation methods, and phases as well as user emotion, user experience and user performance and measuring techniques of them.

Literature review that is displayed in this part of thesis shows there is a worth on investigating user emotion, performance and perceived usability on SIS.

The data picked up from these examinations in regards to the effect of user experiences and usability can be utilized to guide educational institutions' designers and engineers in creating substance and sites that suit the diverse needs of users. The long- term advantage will be better composed practices and better help for the requirements and inclinations of users.

3. TRANSLATION OF THE EMOTION WORD PROMPT LIST INTO TURKISH, THE EWPL-TR

3.1. Introduction

User experience (UX) is a concept universally used in HCI, both in research and practice. Previous examination (Harrison, 2009; Petrie & Harrison, 2009) demonstrated that a few people locate the emotional verbally process hard to do. Users frequently report that it is hard to express their emotions about a system. Keeping in mind the end goal to conquer this issue, it has been investigated the usefulness of giving users an Emotion Word Prompt List (EWPL) that is a rundown of feeling words regularly utilized as a part of depicting websites (Petrie & Harrison, 2009).

EWPL was one of two procedures that Petrie and Precious (2010) utilized as a part of their examinations to create basic yet productive techniques for acquiring UX of sites and other interactive technologies. It was discovered that the usage of EWPL causes users to express their sensitive responses to sites.

EWPL can be adequately used as a rating scale method to be finished after interaction with a site, which is an extremely productive strategy for estimating emotional response to a site. The EWPL questionnaire has been conducted in the literature successfully (Seneler, 2014). Seneler (2014) found that positive feelings of members lead high usability rating of a site and negative feelings lead low usability rating for the site. Along these lines, concentrating on users' emotional responses can possibly raise sites' usability appraisals.

EWPL is quick, understandable and practical scale for assessing user emotion. In EWPL, participants' actual emotional responses to the site were determined using a 16 item scale of emotion words generally used to depict reactions to sites. Participants rated each

emotion word in response to the question “To what extent did the website make you feel each of the following?” on a 5 point scale (from 1 = “Not at all” to 5 = “Very”). EWPL includes 9 positive, 6 negative and 1 ambiguous word (See Table B.1 in Appendix for the EWPL).

The EWPL questionnaire was developed in English. However, the author and supervisor of this thesis have been unsuccessful to find a Turkish version of the EWPL (a fact validated by EWPL’s main author, Professor Dr. Helen Petrie). Therefore, to make it convenient for users and researchers from Turkey, EWPL was translated into the Turkish language with acceptance of Professor Dr. Helen Petrie and its reliability and validity studies have been done.

3.2. Methodology

3.2.1. Translation method of the EWPL into Turkish

With approval from its main author, Professor Dr. Helen Petrie, the EWPL was translated into Turkish.

Keeping in mind the end goal to approve interpretations and to diminish the dangers that can be confronted while making an interpretation of starting with one language then onto the next, two distinctive interpretation strategies were used. Notwithstanding most investigations that have made an interpretation of scales into different dialects have connected one of every interpretation methods amid the interpretation procedure (Isemonger and Sheppard, 2007), the use of both multiple forward and back-interpretation systems in this research anticipated poor translations and empowered interpretations to be crosschecked. To make an interpretation of the EWPL into Turkish, four interpreters who are native speakers of Turkish and advanced speakers of English were utilized. These four interpreters will be alluded to as Translator1, Translator2, Translator3, and Translator4 in this content.

In the first stage, a multiple forward interpretation method was used. A multiple forward interpretation method is the interpretation of a document from the source language into the target language separately by a number of translators (Maxwell, 1996). Translator1 and Translator2 offered two independent interpretations. Then two people as a native speaker of Turkish and advanced English speaker analyzed these interpretations on an item-to-item basis so as to recognize any changes in meaning. Then, Translator3 was asked to translate only the dissimilar parts of the first two interpretations. Later, the works of full three translators were checked and these efforts created an overall first translation (See Figure 3.1).

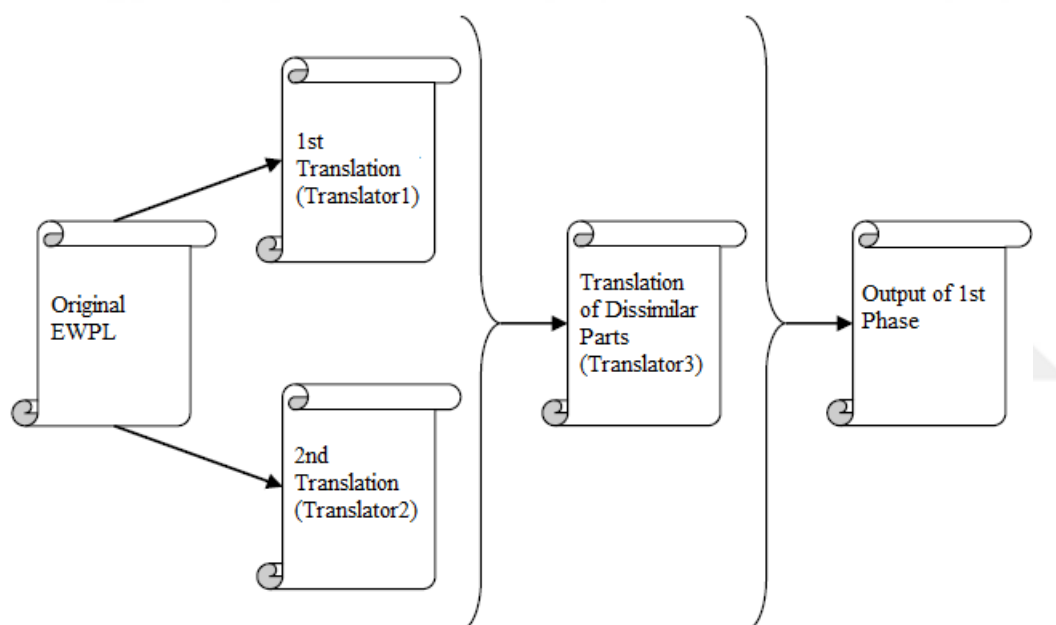


Figure 3.1. : First Step of the Interpretation Method

In the second phase, a back-interpretation method was used, that is a interpretation of a document that has been already translated into a target language back into the origin language (Maxwell, 1996). Translator4 was requested to convert the output of first phase (the overall first interpretation of the EWPL) back into English (See Figure 3.2).

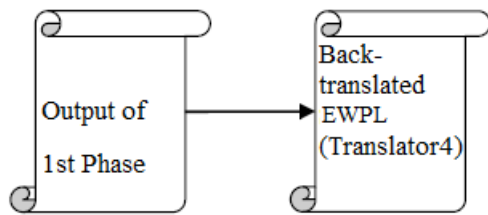


Figure 3.2. : Second Step of the Interpretation Method

Figure 3.3., below, shows the third step of EWPL interpretation process. In the last step of the interpretation process, the original EWPL and the back-interpreted EWPL were examined in contrast. Essential modifications were carried out and the Turkish version of EWPL was finished (See Table B.2 in Appendix for the EWPL-TR).

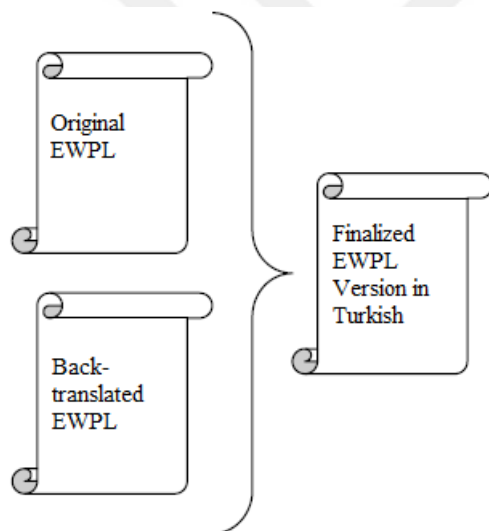


Figure 3.3. : The Third Step of the Interpretation Method

3.2.2. Reliability and validity study of the Turkish version of the EWPL (EWPL-TR)

3.2.2.1. Method

After the translation process has ended, the scale was conducted to study reliability and validation of it. Before putting the scale into practice, a pilot study was carried out with six undergraduate students from a Turkish University to determine unrecognized and unclear

points and to get any feedback. After that, the data were gathered by a two-part survey. In the first part of the survey, participants were asked to give answers on the EWPL-TR. The questions in this section are of the Likert type of 5 (1 = "Not at all" and 5 = "Very"). In the second part of the survey, demographic information for instance age, gender, department plus questions that determine the duration of web usage were included.

In the data collection process, an electronic version of the questionnaire was prepared so that it was applied online. The address of the electronic questionnaire is sent by e-mail to the students and it was also implemented online in several lectures. The data collection process lasted four weeks in total.

3.2.2.2. Participants

The scale was applied to 324 undergraduate students from a Turkish University. As it is illustrated in the following table (See Table 3.1.), 50% of the participants are female and 50% of the participants are male. Almost all participants' ages are between 18-25 years.

Table 3.1. : *Demographic Information of the Working Group*

		f	%
Gender	Female	162	50
	Male	162	50
Age	18	5	1.5
	19	31	9.5
	20	61	18.8
	21	62	19.1
	22	76	23.4
	23	38	11.7
	24	23	7.0
	25	15	4.6
	26	5	1.5
	27	4	1.2
	28	3	0.9
	32	1	0.3
Total		324	100

The average age of the participants is 20. The participants mentioned that 92.9% of them use the web every day. A large majority of participants (70.4%) use computers over 10 years. Almost half of the participants (55.2%) stated that their confidence about computer use could change depending on the given task. Furthermore, 40.1% of the participants expressed that they are confident about using computers. Only one participant mentioned that s/he does not feel confidence and trust about her/his computer usage at all (See Table 3.2).

Table 3.2. : *Web Usage, Computer Usage and Confidence Statistics of the Participants*

	Items	f	%
Web usage	Everyday	301	92.9
	Several times a week	22	6.7
	Several times a month	1	0.3
Total		324	100
Computer usage	Between three and five years	6	1.8
	Between five and seven years	22	6.7
	Between seven and ten years	68	20.9
	Ten years and over	228	70.4
Total		324	100
Computer usage confidence	Do not trust	1	0.3
	Usually, need help	14	4.3
	Depends on the given task	179	55.2
	Trust	130	40.1
Total		324	100

3.2.2.3. *Item analysis*

Item analysis utilizes statistics and master judgment to assess tests in view of the nature of individual things, item sets, and whole arrangements of items, and also the relationship of every item to different items. It investigates the execution of items considered independently either in connection to some outer paradigm or in connection to the rest of the items on the test (Thompson & Levitov, 1985). The items should be extracted from scale if

the relation of one item to the other items is below 0.30 (Buyukozturk, Cakmak, Akgun, Karadeniz & Demirel, 2017). Item-total statistics of the EWPL-TR are given in the following table (See Table 3.3).

Table 3.3. : *Item-total Statistics of the EWPL-TR*

COMPONENTS	Scale Mean if Component Removed	Scale Variance if Item Removed	Corrected Item-Total Correlation	Alpha Value if Component Removed
Item1	33.966	124.708	0.743	0.885
*Item2	33.969	123.776	0.639	0.888
*Item3	33.982	123.802	0.652	0.888
Item4	34.198	127.496	0.614	0.889
*Item5	33.679	129.302	0.437	0.896
Item6	34.756	131.405	0.515	0.893
Item7	34.466	130.113	0.465	0.894
*Item8	34.040	122.107	0.682	0.886
*Item9	34.179	123.559	0.604	0.889
Item10	34.355	125.388	0.771	0.885
Item11	34.244	126.216	0.674	0.887
Item12	34.225	125.890	0.633	0.888
Item13	34.346	125.081	0.762	0.885
Item14	34.386	126.362	0.668	0.888
Item15	33.991	146.003	-0.124	0.916
*Item 16	33.608	124.666	0.520	0.893
*Rotated				

Since the relation of one item to the other items should not be below 0.30, each item that is lower than this value was subtracted from the scale (subtraction was started from the item that has the lowest value). When each item was subtracted from the scale, the item analysis was repeated. This process has been continued until the material has come to a level below 0.30. The final state of the scale is given in the coming table (See Table 3.4). The table

shows the results after the extraction of the item "Item15" from the scale. In the questionnaire form, there are 16 questions related to EWPL. As a result of the item analysis, the number of questions was reduced to 15.

Table 3.4. : *Item-total statistics (after component extraction)*

COMPONENTS	Scale Mean if Component Deleted	Scale Variance if Component Deleted	Corrected Component-Total Correlation	Alpha Value if Item Deleted
Item1	31.531	126.652	0.748	0.907
*Item2	31.534	125.277	0.659	0.910
*Item3	31.546	125.357	0.672	0.909
Item4	31.762	129.761	0.607	0.912
*Item5	31.244	130.730	0.461	0.917
Item6	32.321	133.748	0.505	0.914
Item7	32.031	132.884	0.439	0.917
*Item8	31.605	123.577	0.703	0.908
*Item9	31.744	124.971	0.627	0.911
Item10	31.920	127.653	0.762	0.907
Item11	31.809	128.675	0.658	0.910
Item12	31.790	128.234	0.622	0.911
Item13	31.911	127.141	0.762	0.907
Item14	31.951	128.716	0.656	0.910
*Item16	31.173	125.778	0.552	0.914
*Rotated.				

3.2.2.4. Reliability analysis

Reliability was assessed using coefficient alpha (Cronbach, 1951) and it is generally accepted method for measuring reliability (Sauro and Lewis, 2012). It was determined that the reliability level of the EWPL-TR consisting of 15 items was at a high level which is 0.9 (See Table 3.5).

Table 3.5. : *Reliability Coefficients*

Cronbach's Alpha	N of Items	N of Cases
0.916	15	324

3.2.2.5. Assessment of suitability of data for factor analysis

The Kaiser-Meyer-Olkin (KMO) and Bartlett's test can be used to determine whether the data obtained from the study group is consistent with the explanatory factor analysis (EFA) (Buyukozturk, 2010; Cokluk, Sekercioglu & Buyukozturk, 2012; Karagoz & Kosterelioglu, 2008). As a result of the KMO test, it is interpreted that factor analysis cannot be continued if the KMO value is lower than 0.5 (Cokluk et al., 2012). The KMO value for this study was 0.9 that is quite adequate for research sample. Showing the suitability of the data for factor analysis, Bartlett's test result was also significant ($\chi^2 = 2668.006$, $p = 0.000$) (See Table 3.6).

Table 3.6. : *KMO and Bartlett test results*

KMO Test		0.936
Bartlett's Test	Chi-square	2668.006
	df	105
	p.	0.000

3.2.2.6. Construct validity

For evaluating construct validity, the results of both EFA and CFA were declared for the items of the EWPL-TR. In order to measure construct validity, LISREL Statistic Solution has been utilized. Based on the findings, the principal component method and varimax rotation were applied as EFA of the 15-item of the EWPL-TR. As shown in the following table, the 15-item EWPL-TR appears to be a two-factor structure and give reason for 60.0%

of the total two-factor variance. The first factor identified 31.6% of the total variance and the second factor identified 28.4% of total variance. A two-factor structure emerged that explains 60.0% of the total variance resulting from factor analysis with an eigenvalue greater than 1 (See Table 3.7).

Table 3.7.: *Total variance explained*

Components	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	% Of the Variance	Cumulative %	Total	% Of the Variance	Cumulative %	Total	% Of the Variance	Cumulative %
1	7.194	47.960	47.960	7.194	47.960	47.960	4.733	31.553	31.553
2	1.805	12.033	59.993	1.805	12.033	59.993	4.266	28.440	59.993
3	0.850	5.667	65.660						
4	0.645	4.300	69.960						
5	0.631	4.209	74.168						
6	0.561	3.737	77.905						
7	0.547	3.646	81.551						
8	0.517	3.446	84.997						
9	0.405	2.698	87.696						
10	0.389	2.591	90.287						
11	0.356	2.372	92.659						
12	0.328	2.187	94.846						
13	0.283	1.886	96.732						
14	0.249	1.661	98.393						
16	0.241	1.607	100.000						

When the scree plot test graph is examined (See Figure 3.4.), it is seen that limiting the factor number to two is sufficient. When the distribution of the items according to the factors

after the varimax rotation method is examined, it is seen that all the materials provide logical integrity in terms of the factor structures.

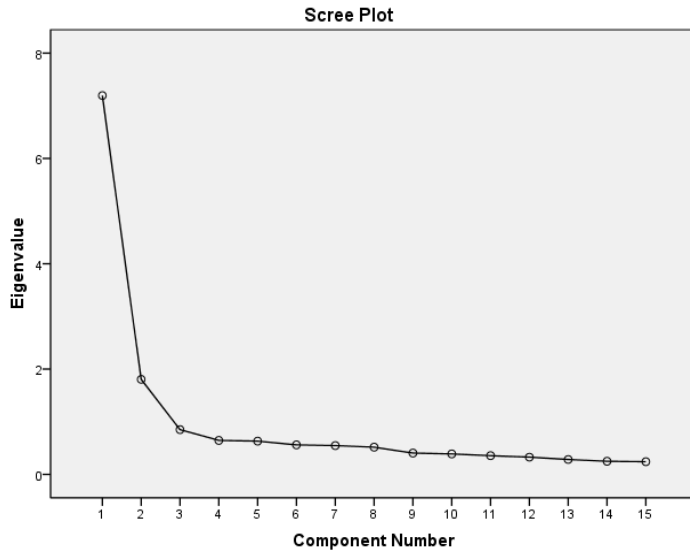


Figure 3.4. : Scree Plot Test Graph

According to the factor loadings which are shown in the following table (See Table 3.8.), since the distances between the loads of the factors in which the items are collected must be at least 10% of the distance between them, there is no item which does not comply with this rule. As a general rule, regardless of the sign value, the load value higher than 0.60 is high, and the load value between 0.30 and 0.59 can be defined as moderate values (Büyüköztürk, 2002).

Table 3.8. : *Varimax-rotated two-factor result for the EWPL-TR*

EWPL- TR	Factors	
	1	2
EM1	0.538	
EM6	0.582	
EM4	0.658	
EM13	0.711	
EM7	0.714	
EM14	0.720	
EM10	0.756	
EM12	0.790	
EM11	0.804	
EM2		0.726
EM3		0.783
EM5		0.652
EM8		0.782
EM9		0.816
EM16		0.668

When the items belonging to the factors are examined; nine expressions were collected under the first factor, and these expressions and load values and other statistical values are shown in the following table (See Table 3.9). This factor can be named as "Positive EWPL-TR factor". The factor loads of the positive EWPL-TR factor were determined to be between 0.53 and 0.80. The positive EWPL-TR factor has a high reliability, which is 0.91.

Table 3.9. : *Positive EWPL-TR Factor*

Factor 1: Positive EWPL-TR Factor	Factor Load	Factor Reliability
EM1	0.538	0.905
EM6	0.582	
EM4	0.658	
EM13	0.711	
EM7	0.714	
EM14	0.720	
EM10	0.756	
EM12	0.790	
EM11	0.804	

Six expressions were collected under the second factor, and these expressions, load values, and other statistical values are shown in the following table (See Table 3.10). This factor can be named as "Negative EWPL-TR factor". The factor loads of the negative EWPL-TR factor were determined to be between 0.65 and 0.82. The negative EWPL-TR factor has a high reliability, which is 0.86.

Table 3.10. : *Negative EWPL-TR Factor*

Factor 2: Negative EWPL-TR Factor	Factor Load	Factor Reliability
EM15	0.668	0.864
EM5	0.652	
EM2	0.726	
EM3	0.783	
EM8	0.782	
EM9	0.816	

3.2.2.7. CFA results

The accuracy of the two-dimensional factorial structure based on the results of EFA was tested by CFA. Since there is no single statistical significance test used to calculate the fitness of the model generated using the obtained data, the fact that many measurements are considered simultaneously in the process of evaluating the model in the study has been taken

into consideration. In the CFA, different indices were used to evaluate the fitness of a model, and the most commonly used ones are; the root mean square error of approximation (RMSEA) and comparative fit index (CFI) (Tabachnick & Fidell, 2007). In this study, the values of chi-square (χ^2), RMSEA, CFI, the goodness of fit index, adjusted good fit index (AGFI) and normed fit index (NFI) were evaluated on the basis of.

When performing the CFA analysis of EWPL-TR, the subscales of the scale were called as positive dimension and negative dimension. Necessary modifications have been carried out and the EWPL-TR scale has been modified from EM11 to EM7 and from EM14 to EM13 (See Figure 3.5). As shown in the following figure, EM10 is the most important item with a coefficient of 0.86 in the positive dimension whereas EM8 is the most important item with a coefficient of 0.82 in the negative dimension. In addition, the relationship between positive and negative dimension is 0.7, and the relationship between them is significant ($p=0.000$). The results of the CFA analysis in which the subscales of the EWPL-TR are included are given in the following table (See Table 3.11).

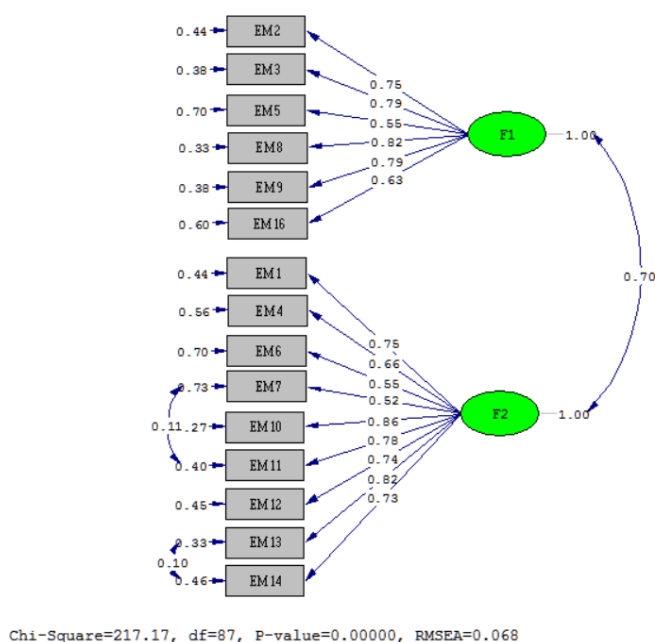


Figure 3.5. : CFA Results for EWPL-TR

Table 3.11. : *EWPL-TR model results*

Factor/Item	Standardized Loads	t-value	R²
Positive Dimension			
EM1	0.75	15.42	0.56
EM4	0.66	13.04	0.44
EM6	0.55	10.36	0.30
EM7	0.52	9.61	0.27
EM10	0.73	18.77	0.73
EM11	0.78	16.17	0.60
EM12	0.74	15.10	0.55
EM13	0.82	17.59	0.67
EM14	0.73	14.83	0.054
Negative Dimension			
EM2	0.75	15.12	0.57
EM3	0.79	16.35	0.62
EM5	0.55	10.23	0.51
EM8	0.82	17.28	0.67
EM9	0.79	16.35	0.62
EM16	0.63	12.11	0.40

It is possible to say that the compliance criteria for the CFA analysis of EWPL-TR are among the acceptable limits. Other than these criteria, $\chi^2 (87) = 217.17$; $\chi^2 / df = 2.50 < 3$ is another indicator used to determine the model suitability and is another indicator that the model is perfectly in terms of statistics (See Table 3.12). The table shows that the RMSEA value of the developed scale is in acceptable limits (RMSEA = 0.07). Furthermore, the SRMR and GFI values are within the acceptable fit index (SRMR = 0.05) and (GFI = 0.9). Besides the AGFI value is within the fit index (AGFI = 0.9). As a result of the CFA analysis, the items were confirmed to have related factors at 95% confidence level ($p < 0.05$) and that the compliance indices were within the acceptable compliance values and that the model consistency was within the acceptable values ($\chi^2 / df = 2.50 < 3$).

Table 3.12. : *Values of Compliance Criteria for the EWPL Scale Cfa Model.*

Compliance Criteria	Good Fit	Acceptable Fit	Values Of Developed Scale (Current Study)
RMSEA	$0 < \text{RMSEA} < 0.05$	$0.05 \leq \text{RMSEA} \leq 0.10$	0.068
SRMR	$0 \leq \text{SRMR} < 0.05$	$0.05 \leq \text{SRMR} \leq 0.10$	0.054
GFI	$0.95 \leq \text{GFI} \leq 1$	$0.90 \leq \text{GFI} \leq 0.95$	0.920
AGFI	$0.90 \leq \text{AGFI} \leq 1$	$0.85 \leq \text{AGFI} \leq 0.90$	0.890
RMSEA: Root Mean Square Error of Approximation, SRMR: Standardized Root Mean Square Residual, GFI: Goodness of Fit Index, AGFI: Adjusted Goodness of Fit Index (Schermelleh-Engel, Moosbrugger, and Müller, 2003)			

3.3. Findings

To make an official Turkish version of the EWPL suitable for users and researchers from Turkey, the EWPL was interpreted into the Turkish language with approval of Professor Dr. Helen Petrie by using professional translation methods. In this study, the use of both multiple forward and back-translation methods blocked poor interpretations and allowed interpretations to be cross-checked. In order to translate the EWPL into Turkish, four interpreters who are national speakers of Turkish and native speakers of English were used. With this effort, a definite translation has been performed and the EWPL-TR was structured.

Furthermore, the scale was applied to 324 undergraduate students from a Turkish University. It was assessed for its reliability and validity and was found to have good reliability and validity. The reliability level of the EWPL-TR scale consisting of 15 items was at a high level, which is 0.9. The KMO value for this study was 0.9 that is quite adequate for research sample. Showing the suitability of the data for factor analysis, Bartlett's test result was also significant ($\chi^2 = 2668.006$, $p = 0.000$).

For assessing construct validity, the results of both EFA and CFA were demonstrated for the items of the EWPL-TR. As a result of the item analysis, the number of questions was

reduced to 15, so as to get better outcomes. Based on the findings, the principal component method and varimax rotation were applied as EFA of the 15-item EWPL-TR scale. The 15-item EWPL-TR appears to be a two-factor structure and gives reason for 60.0% of the total two-factor variance. The first factor identified 31.5% of the total variance and the second factor identified 28.4% of total variance. A two-factor structure emerged that explains 60.0% of the total variance resulting from factor analysis with an eigenvalue greater than 1.

When the distribution of the items according to the factors after the varimax rotation method is examined, it is seen that all the materials provide logical integrity in terms of the factor structures. When the items belonging to the factors are examined, two factors emerged. This finding is in line with studies in literature (Seneler, 2014). The compliance criteria for the CFA analysis of the EWPL-TR are among the acceptable limits.

3.4. Discussion

In this study, the EWPL (Petrie & Harrison, 2009), which was developed to measure user emotion with systems, was adapted to Turkish. The current study has resulted in the development and validation of the EWPL-TR for usability studies in Turkey. The multi-stage interpretation process included the steps of initial translation, expert review, and back-translation. Psychometric evaluation of the EWPL-TR illustrated a satisfactory level of reliability. EWPL-TR scale appears to be a two-factor structure that is in line with the literature. It is possible to say that the compliance criteria for the CFA analysis of the EWPL-TR are among the acceptable limits.

Measuring emotion which is very important for both the usability studies and psychological experiments have been very limited in Turkey. Therefore, Turkish people group have an awesome requirement for substantial and solid instruments and tools to quantify users' feeling of the extensive variety of products and services. The aftereffects of this

examination demonstrated that the EWPL-TR is a valid and reliable instrument for estimating user emotion.

As a conclusion, now the EWPL is available with Turkish users. Turkish academicians and researchers who wish to attempt look into with Turkish members identified with the EWPL can use EWPL-TR.



4. TRANSLATION OF THE SUS INTO TURKISH, THE SUS-TR

4.1. Introduction

The increase of developing and using complex technologies causes the need for measuring the quality of users' experiences when they are interacting with a system. Therefore, researches on user experience and usability still get attention to gain a better understanding of these issues. There are lots of studies that showed usable systems' positive effects on users' experiences (Forlizzi & Battarbee, 2004; Kampf & Payne, 2000; Thüring & Mahlke, 2007). Nielsen (2012) stated that if a website is not easy to use, people leave. In today's world, people can easily switch to another system since there are lots of alternatives. Thus, it is highly critical to create and develop usable and effective systems. In addition, usability also can affect users' preferences on systems. According to Jordan (1998), if a system is not usable this can cause problems of a varying degree of asperity. Besides, usability has financial associations for the commercial and industrial world (Mack & Sharples, 2009). Mack and Sharples (2009) studied the importance of usability on users' system preference related to system's attributes. They found that usability is important on system preference. Thus, usability evaluation and its measurement are extremely important both during the process of system development and after the system was released.

Since systems are developed for users, usability mainly focuses on understanding users. That's why, if developers desire to develop a useful system, the understanding of potential users is vital. Hence, the last approach that includes user-based (also empirical) methods focus on observing how users are interacting with systems or allow users to express their perceptions about systems' usability. According to several studies, most apparent way to understand about the usability of the tested system is asking to the participants about their experiences with it (Tullis & Albert, 2013) and this can be fulfilled in a structured way by

using a questionnaire (Macleod, 1994; Tullis & Albert, 2013). Questionnaires support researchers to understand potential users clearly (Rubin & Chisnell, 2008). In a broad way, they are appropriate methods for interpreting how users use systems and what features they specifically like or dislike (Nielsen, 1993). For these reasons, questionnaires are greatly practical methods for assessing usability.

A variety of surveys have been used and reported in the literature for measuring perceived usability of information systems such as System Usability Scale (SUS) (Brooke, 1996), Questionnaire for User Interface Satisfaction (QUIS) (Harper & Norman, 1998) and Computer System Usability Questionnaire (CSUQ) (Lewis, 1995). The SUS, has been widely used in the literature (Barnum, 2011; Brooke, 2013; Finstad, 2006; Tullis & Albert, 2013, Tullis & Stetson, 2004). The SUS was created in the 1980s by Dr. John Brooke at Digital Equipment Co. Ltd. to give a universal view of subjective evaluations of usability (Brooke, 1996). It is simple and practical scale as a result is widely accepted and used for assessing the usability of diverse types of systems.

The SUS is a 10-item survey (See Table B.3. in Appendix for the first form of the SUS) in which respondents exhibit their level of concurrence with every item on a scale from 1 (strongly disagree) to 5 (strongly agree). The odd numbered things (1,3,5,7 and 9) have positive-tone words and the even-numbered things (2,4,6,8 and 10) have negative-tone words. For odd-numbered things, the score commitment is the scale position less 1 and for even-numbered things, the score commitment is 5 less the scale position. Next, to get the general estimation of SUS, the aggregate of the commitments ought to be duplicated by 2.5. The calculation of SUS scores to a scale that can range from 0 to 100 makes it easier for usability practitioners to connect (Brooke, 2013). Brooke (1996) stated that higher scores indicate better system usability.

Nunnally (1978) stated that standardized usability questionnaires offer many advantages, which are objectivity, replicability, quantification, economy, communication and scientific generalization to practitioners. The SUS is potent, notorious and its use is well-established with more than 1200 publications (Brooke, 2013). There are few direct comparisons of the various standardized usability questionnaires (Sauro & Lewis, 2012). Moreover, its comparison with other usability questionnaires have been performed (Tullis & Stetson; 2004). Tullis and Stetson (2004) conducted a comparison of questionnaires for assessing website usability study. Five questionnaires (SUS, QUIS, CSUQ, a two variant of Microsoft's Product Reaction Cards) were compared with 123 participants. The SUS produced among the most reliable results across sample sizes. Besides, the SUS was the only questionnaire that has questions all address distinct aspects of the user's reaction to the website as a whole.

The SUS questionnaire was developed in English. There are studies that showed participants respond better to scales in their own language (Bahrnick, Hall, Goggin, Bahrnick, and Berger, 1994; Delgado, Guerrero, Goggin and Ellis, 1999). The SUS is previously translated European Portuguese, Persian and Slovene languages (Blažica and Lewis, 2015; Dianat, Ghanbari, and AsghariJafarabadi, 2014; Martins, Rosa, Queirós, Silva, and Rocha, 2015). There are also various unofficial translations of the SUS. Its reliability and validity studies are also carried out. Bangor, Kortum and Miller (2008) described the results of 2324 SUS surveys from 206 usability tests collected over a 10-year period. In that study, it was found that the SUS was highly reliable (Cronbach's $\alpha = 0.91$) and useful over a wide range of interface types. A number of researchers have noted the tendency for positive-tone and negative-tone items of the SUS to load on separate factors (Barnette, 2000; Davis, 1989; Pilotte & Gable, 1990; Sauro & Lewis, 2011; Schmitt & Stuitts, 1985; Schriesheim & Hill, 1981; Wong, Rindfleisch, & Burroughs, 2003).

4.1.1. The Distinction between the SUS-TR and Kadirhan, Gül and Battal's Study

Despite the fact that various interpretations into different dialects exist as specified at over, the authors of this thesis have neglected to locate a Turkish rendition of the SUS (a reality affirmed by SUS's fundamental creator, Dr. John Brooke). Over the span of leading this work, it was found that the SUS had just been converted into Turkish. Nonetheless, it has been discovered numerous issues with that interpretation. Not at all like the past interpretation, it has been utilized proficient interpretation systems amid the improvement of the survey. The SUS was converted into Turkish with the assistance of academics in Kadirhan, Gül and Battal's study though four unique interpreters took an interest in the advancement procedure of the SUS-TR. Also, both multiple forward and back-interpretation procedures were utilized as a part of the development procedure of the SUS-TR to counteract poor interpretations and empower interpretations to be crosschecked.

In addition, it has been established the reliability and validity of the questionnaire by receiving permission was obtained from the main author of the SUS who is Dr. John Brooke. Therefore this research programme contributes useful Turkish version of the SUS. Therefore, to make an official Turkish version of the SUS (SUS-TR) suitable for users from Turkey, it was translated into the Turkish language with permission of Dr. John Brokee by using professional translation techniques. Furthermore, the SUS-TR was studied for its reliability and validity.

For the reasons that are mentioned at above, the SUS was translated into Turkish and the studies on reliability and validity of the questionnaire were performed.

4.2. Translation Of The SUS Into Turkish

4.2.1. Translation Method Of The SUS Into Turkish

With approval from its main author, Dr. John Brooke, the SUS was translated into Turkish. Keeping in mind the end goal to approve interpretations and to diminish the dangers that can be confronted while making an interpretation of starting with one language then onto the next, two distinctive interpretation strategies were used. Notwithstanding most investigations that have made an interpretation of scales into different dialects have connected one of every interpretation methods amid the interpretation procedure (Isemonger & Sheppard, 2007), the use of both multiple forward and back-interpretation systems in this research anticipated poor translations and empowered interpretations to be crosschecked. To make an interpretation of the SUS into Turkish, four interpreters who are native speakers of Turkish and advanced speakers of English were utilized. These four interpreters will be alluded to as Translator1, Translator2, Translator3, and Translator4 in this content.

In the first stage, a multiple forward interpretation method was used. A multiple forward interpretation technique is the understanding of a report from the source dialect into the objective dialect independently by various interpreters (Maxwell, 1996). Translator1 and Translator2 offered two free interpretations. At that point two individuals as a local speaker of Turkish and advanced English speaker dissected these interpretations on a item to-item premise to perceive any adjustments in significance. At that point, Translator3 was requested to decipher just the different parts of the initial two interpretations. Afterward, crafted by full three interpreters were checked and these endeavors made a general first interpretation (See Figure 4.1).

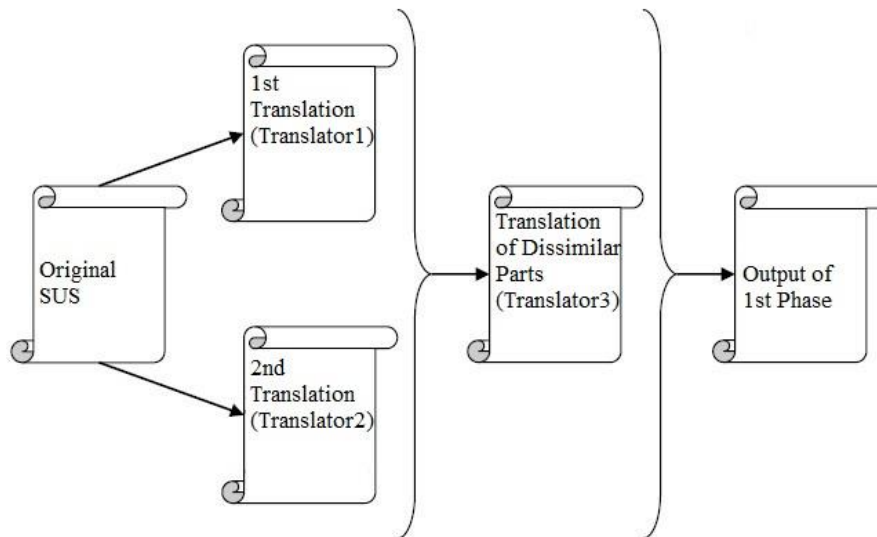


Figure 4.1. : The first step of the Interpretation Method

In the second phase, a back-interpretation technique was used, that is a interpretation of a document that has been already interpreted into a target language back into the origin language (Maxwell, 1996). Translator4 was requested to convert the output of first phase (the overall first interpretation of the SUS) back into English (See Figure 4.2).

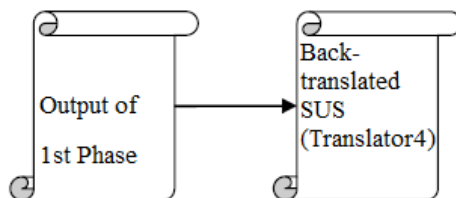


Figure 4.2. : Second Step of the Interpretation Method

In the third and last step of the translation process, the original SUS and the back-interpreted SUS were compared (See Figure 4.3). Appropriate modifications were made and the Turkish version of SUS was concluded (See Table B.4 in Appendix for the SUS-TR).

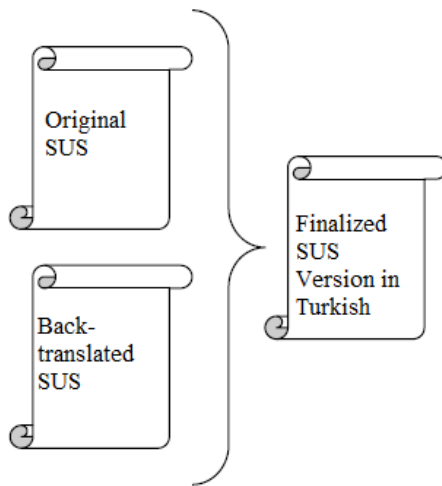


Figure 4.3. : Third Step of the Interpretation Method

4.2.2. Reliability and Validity Study of the SUS-TR

4.2.2.1. Method

After the translation process has ended, the scale was conducted to study reliability and validation of it. Before putting the scale into practice, a pilot study was carried out with six undergraduate students from a Turkish University to determine unrecognized and unclear points and to get any feedback. After that, the data were collected by a two-part questionnaire. In the first part of the questionnaire, participants were asked to give answers on the SUS-TR. In the second part of the questionnaire, demographic information such as age, gender, department, and questions that determine the duration of web usage were included.

In the data collection process, an electronic version of the questionnaire was prepared so that it was applied online. The address of the electronic questionnaire is sent by e-mail to the students and it was also implemented online in several lectures. The data collection process lasted four weeks in total.

4.2.2.2. Participants

The scale was applied to 324 undergraduate students from a Turkish University. As it is demonstrated in the following table (See Table 4.1.), 50% of the participants are female and 50% of the participants are male. Almost all participants' ages are between 18-25 years.

Table 4.1. : *Demographic Information of the Working Group*

		f	%
Gender	Female	162	50
	Male	162	50
Age	18	5	1.5
	19	31	9.5
	20	61	18.8
	21	62	19.1
	22	76	23.4
	23	38	11.7
	24	23	7.0
	25	15	4.6
	26	5	1.5
	27	4	1.2
	28	3	0.9
	32	1	0.3
Total		324	100

The average age of the participants is 20. The participants mentioned that 92.9% of them use the web every day. A large majority of participants (70.4%) use computers over 10 years. Almost half of the participants (55.2%) stated that their confidence about computer use could change depending on the given task. Furthermore, 40.1% of the participants expressed that they are confident about using computers. Only one participant mentioned that s/he does not feel confidence and trust about her/his computer usage at all (See Table 4.2).

Table 4.2. : *Web Usage, Computer Usage and Confidence Statistics of the Participants*

	Items	f	%
Web usage	Everyday	301	92.9
	Several times a week	22	6.7
	Several times a month	1	0.3
Total		324	100
Computer usage	Between three and five years	6	1.8
	Between five and seven years	22	6.7
	Between seven and Ten years	68	20.9
	Between ten years and over	228	70.4
Total		324	100
Computer usage confidence	Do not trust	1	0.3
	Usually, need help	14	4.3
	Depends on the given task	179	55.2
	Trust	130	40.1
Total		324	100

4.2.2.3. Item analysis

Item analysis utilizes statistics and master judgment to assess tests in view of the nature of individual things, item sets, and whole arrangements of items, and also the relationship of every item to different items. It investigates the execution of items considered independently either in connection to some outer paradigm or in connection to the rest of the items on the test (Thompson & Levitov, 1985). The items should be extracted from scale if the relation of one item to the other items is below 0.30 (Buyukozturk et al, 2017). Item-total statistics of the EWPL-TR are given in the coming table (See Table 4.3).

Table 4.3. : *Item-total Statistics of the SUS-TR*

COMPONENTS	Scale Mean if Component Removed	Scale Variance if Component Removed	Corrected Component- Total Correlation	Alpha Value if Item Deleted
*Item 4	29.417	59.160	0.397	0.841
Item 3	28.247	55.382	0.561	0.826
Item 7	28.164	54.986	0.587	0.823
*Item 10	29.056	56.926	0.515	0.830
*Item 6	28.321	54.417	0.665	0.816
*Item 8	28.531	56.423	0.583	0.824
*Item 2	28.500	53.594	0.653	0.816
Item 5	27.546	59.394	0.548	0.829
Item 9	27.907	58.555	0.516	0.830
Item1	27.617	60.807	0.371	0.842
*Rotated				

4.2.2.4. Reliability analysis

Reliability was assessed using coefficient alpha (Cronbach, 1951) and it is generally accepted method for measuring reliability (Sauro & Lewis, 2012). It was determined that the reliability level of the SUS-TR scale consisting of 10 items was at a high level which is 0.8 (See Table 4.4).

Table 4.4. : *Reliability coefficients*

Cronbach's Alpha	N of Items	N of Cases
0.84	10	324

4.2.2.5. Assessment of appropriateness of data for factor analysis

The Kaiser-Meyer-Olkin (KMO) and Barlett's test can be used to determine whether the data obtained from the study group is consistent with the explanatory factor analysis (ECA) As a result of the KMO test, it is interpreted that factor analysis cannot be continued if the value is lower than 0.50. The KMO value for this study was 0.87 that is quite adequate for research sample. Showing the suitability of the data for factor analysis, the Bartlett test result was also significant ($\chi^2 = 1072.87$, $p = 0.000$) (See Table 4.5).

Table 4.5. : *KMO and Bartlett's Test Results*

Kaiser-Meyer-Olkin Test (KMO)		0.871
Bartlett's Test	Chi-Square	1072.869
	df	45
	p.	0.000

4.2.2.6. Construct validity

For evaluating construct validity, the results of both EFA and CFA were presented for the items of the SUS-TR. In order to measure construct validity, LISREL Statistic Solution has been utilized. Based on the findings, the principal component method and varimax rotation were applied as EFA of the 10-item SUS-TR scale. A two-factor structure emerged that explains 56.7% of the total variance resulting from factor analysis with an eigenvalue greater than 1. The first factor and the second factor revealed 29.5% and 27.3% of the total variance, respectively (See Table 4.6).

Table 4.6. : *Total Variance Explained*

	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	% Of the variance	Cumulative %	Total	% Of the variance	Cumulative %	Total	% Of the variance	Cumulative %
1	4.210	42.096	42.096	4.210	42.096	42.096	2.936	29.365	29.365
2	1.460	14.602	56.698	1.460	14.602	56.698	2.733	27.333	56.698
3	0.809	8.095	64.793						
4	0.674	6.738	71.531						
5	0.582	5.820	77.352						
6	0.541	5.411	82.763						
7	0.509	5.088	87.851						
8	0.447	4.475	92.326						
9	0.443	4.427	96.753						
10	0.325	3.247	100.000						

When the scree plot test graph is examined (See Figure 4.4.), it is seen that limiting the factor number to two is sufficient. When the distribution of the items according to the factors after the varimax rotation method is examined, it is seen that all the materials provide logical integrity in terms of the factor structures.

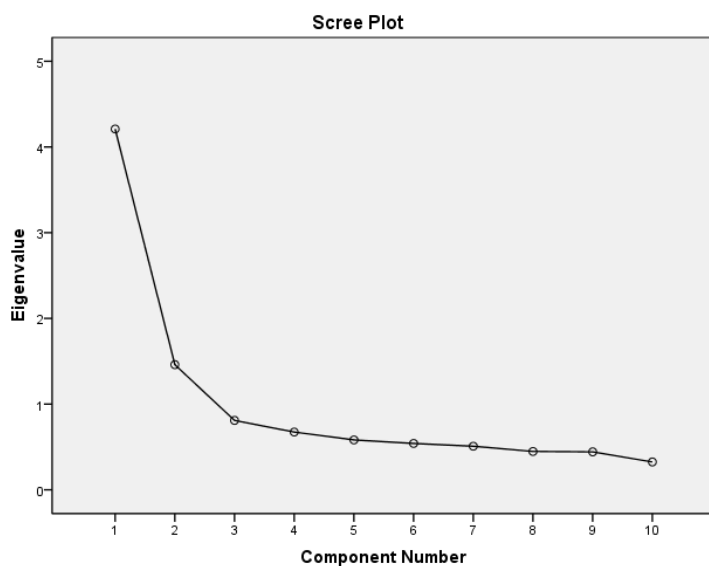


Figure 4.4. : Scree plot test graph

According to the factor loadings which are shown in the following table (See Table 4.7.), since the distances between the loads of the factors in which the items are collected must be at least 10% of the distance between them, there is no item which does not comply with this rule. The Cronbach Alpha value was used to calculate the internal consistency of the factors. If the Cronbach alpha value is over 0.70, it means that the reliability is high (Cronbach, 1951). As shown in the following table, all items have Cronbach values greater than 0.70.

Table 4.7.: *Varimax-rotated Two-factor Solution for the Turkish Version of the System Usability Scale*

Items	Factors	
	1	2
SUS5	0.775	
SUS1	0.746	
SUS7	0.725	
SUS3	0.609	
SUS9	0.605	
SUS10		0.819
SUS4		0.753
SUS8		0.684
SUS6		0.640
SUS2		0.600

When the items belonging to the factors are examined, five expressions were collected under factor 1, and these expressions, load values, and other statistical values are shown in the following table (See Table 4.8). This factor can be named as "Positive SUS-TR Factor". Factor loads of the positive SUS-TR factor ranged from 0.61 to 0.78. Positive SUS-TR factor has a high reliability, which is 0.77.

Table 4.8. : *Positive SUS-TR Factor*

Factor 1: Positive SUS-TR Factor	Factor Load	Factor Reliability
SUS5	0.775	0.769
SUS1	0.746	
SUS7	0.725	
SUS3	0.609	
SUS9	0.605	

Five expressions were collected under factor 2 (See Table 4.9.), and these expressions and load values and other statistical values are shown in the following table. When the collected items of this factor are examined, the factor can be named as "Negative SUS-TR Factor". Factor loads of the negative SUS-TR factor ranged from 0.60 to 0.82. Negative SUS-TR factor has a high reliability, which is 0.81.

Table 4.9. : *Negative SUS-TR Factor*

Factor 2: Negative SUS-TR Factor	Factor Load	Factor Reliability
SUS10	0.819	0.805
SUS4	0.753	
SUS8	0.684	
SUS6	0.640	
SUS2	0.600	

4.2.2.7. CFA results

The accuracy of the two-dimensional factorial structure based on the results of EFA was tested by CFA. Since there is no single statistical significance test used to assess the fitness of the model generated using the obtained data, the fact that many measurements are considered simultaneously in the process of evaluating the model in the study has been taken into consideration. In the CFA, different indices were used to assess the fitness of a model, and the most commonly used ones were; "The Root Mean Square Error of Approximation (RMSEA)" and "Comparative Fit Index (CFI)" (Tabachnick and Fidell, 2007). In this study, the values of chi-square (χ^2), RMSEA, CFI, "Goodness of Fit Index", "Adjusted Good Fit Index (AGFI)" and "Normed Fit Index (NFI)" were evaluated on the basis of.

When performing the CFA analysis of SUS-TR, the subscales of the scale were called as positive dimension and negative dimension. Necessary modifications have been carried out and the SUS-TR scale has been modified from SUS9 to SUS7 and from SUS10 to SUS4. As shown in the following figure (See Figure 4.9.), SUS7 is the most important item with a coefficient of 0.76 in the positive factor whereas SUS6 is the most important item with a coefficient of 0.80 in the negative factor.

In addition, the relationship between positive and negative dimensions is 0.71, and the relationship between them is significant ($p = 0.000$). The results of the CFA analysis in which the subscales of the SUS-TR are included are given in the following table (See Table 4.10).

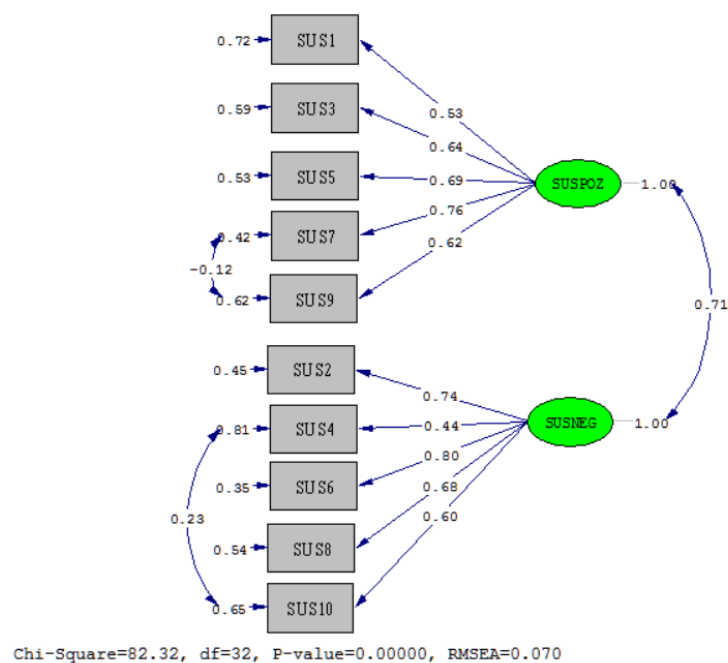


Figure 4.5. : CFA results for SUS-TR scale

Table 4.10. : SUS-TR Scale Model Results

Factor/Item	Standardized Loads	t-value	R ²
Positive dimension			
SUS1	0.53	9.35	0.28
SUS3	0.64	11.85	0.41
SUS5	0.69	12.89	0.47
SUS7	0.76	14.32	0.58
SUS9	0.62	10.75	0.38
Negative dimension			
SUS2	0.74	14.40	0.55
SUS4	0.44	7.62	0.19
SUS6	0.80	15.95	0.65
SUS8	0.68	12.73	0.46
SUS10	0.60	10.84	0.35

It is possible to say that the compliance criteria for the CFA analysis of SUS-TR are among the acceptable limits. Other than these criteria, $\chi^2 (32) = 82.32$, $\chi^2 / df = 2.57 < 3$ is another indicator used to determine the model suitability and is another indicator that the model is perfectly in terms of statistics (See Table 4.11).

Table 4.11. : *Values of compliance criteria for the SUS-TR scale CFA model*

Compliance Criteria	Good Fit	Acceptable Fit	Values of Developed Scale (Current Study)
RMSEA	$0 < \text{RMSEA} < 0.05$	$0.05 \leq \text{RMSEA} \leq 0.10$	0.070
SRMR	$0 \leq \text{SRMR} < 0.05$	$0.05 \leq \text{SRMR} \leq 0.10$	0.052
GFI	$0.95 \leq \text{GFI} \leq 1$	$0.90 \leq \text{GFI} \leq 0.95$	0.950
AGFI	$0.90 \leq \text{AGFI} \leq 1$	$0.85 \leq \text{AGFI} \leq 0.90$	0.920
RMSEA: Root Mean Square Error of Approximation, SRMR: Standardized Root Mean Square Residual, GFI: Goodness of Fit Index, AGFI: Adjusted Goodness of Fit Index (Schermelleh-Engel, Moosbrugger, and Müller, 2003)			

The previous table shows that the RMSEA value of the developed scale is acceptable ($\text{RMSEA} = 0.070$, $0.00 \leq \text{RMSEA} \leq 0.05$). Furthermore, the SRMR value is within the acceptable fit index ($\text{SRMR} = 0.052$, $0.00 \leq \text{SRMR} \leq 0.05$) ($\text{GFI} = 0.950$, $0.95 \leq \text{GFI} \leq 1.00$). Besides, the AGFI value was within the perfect fit index ($\text{AGFI} = 0.920$, $0.90 \leq \text{AGFI} \leq 1.00$). As a result of the CFA analysis, the items were confirmed to have related factors at 95% confidence level ($p < 0.05$) and that the compliance indices were within acceptable values and the model consistency was within acceptable values ($\chi^2 / df = 2.57 < 3$).

4.3. Findings

In order to make an official Turkish version of the SUS suitable for users from Turkey, it was translated into the Turkish language with permission of Dr. John Brooke by using professional interpretation techniques. In this study, the use of both multiple forward

and back-translation techniques prevented poor interpretations and enabled translations to be crosschecked. In order to translate the SUS into Turkish, four translators who are native speakers of Turkish and advanced speakers of English were employed. With this effort, a definite translation has been performed and the SUS-TR was structured.

Furthermore, the scale was applied to 324 undergraduate students from a Turkish University. It was evaluated for its reliability and validity and was found to have good reliability and validity. The reliability level of the SUS-TR scale consisting of 10 items was at a high level, which is 0.84. The KMO value for this study was 0.87 that is quite adequate for research sample. Showing the suitability of the data for factor analysis, the Bartlett test result was also significant ($\chi^2 = 1072.87$, $p = 0.000$).

For evaluating construct validity, the results of both EFA and CFA were introduced for the items of the SUS-TR. Based on the findings, the principal component method and varimax rotation were applied as EFA of the 10-item SUS-TR scale. A two-factor structure emerged that explains 56.7% of the total variance resulting from factor analysis with an eigenvalue greater than 1. The first factor revealed 29.5% of the total variance and the second factor revealed 27.3% of total variance. When the distribution of the items according to the factors after the varimax rotation method is examined, it is seen that all the materials provide logical integrity in terms of the factor structures. When the items belonging to the factors are examined, two factors emerged. This finding is in line with studies in literature (Barnette, 2000; Davis, 1989; Pilotte & Gable, 1990; Sauro & Lewis, 2011; Schmitt & Stuitts, 1985; Schriesheim & Hill, 1981; Wong et al., 2003). The compliance criteria for the CFA analysis of the SUS-TR are among the acceptable limits.

4.5. Discussion

In this study, the SUS, which was developed to measure usability, was adapted to Turkish. The present study has resulted in the development and validation of the SUS-TR for usability studies in Turkey. The multi-stage interpretation process included the steps of initial interpretation, expert review, and back-translation. Psychometric evaluation of the SUS-TR indicated an acceptable level of reliability. SUS-TR scale appears to be a two-factor structure that is in line with the literature. It is possible to say that the compliance criteria for the CFA analysis of the SUS-TR are among the acceptable limits.

Usability researches, which are essential for both the designers of the products and end-users, have been exceptionally restricted in Turkey. The aftereffects of this investigation demonstrated that the SUS-TR is a valid and reliable tool for estimating usability, with psychometric properties reliable with the first English version.

As a conclusion, now the SUS is usable with Turkish users. Turkish researchers who wish to undertake research with Turkish participants related to the SUS can use SUS-TR.

5. INVESTIGATING THE EFFECTS OF DIFFERENT STUDENT GROUPS ON USERS' EMOTION AND PERCEIVED USABILITY OF SIS

This research programme was created to focus on the following research question of the thesis:

How do students' demographic information affect their experience and perceived usability with a SIS?

5.1. Method

Each participant performed a series of short questionnaires: (a) their urgent emotional response to the system (as measured by the EWPL-TR); (b) their perception of the usability of the system (as measured by the SUS-TR); (c) their like/dislike on the SIS (as measured by the Overall Reactions Questionnaire, ORQ).

Thus, there were two groups of dependent variables:

- a) User experience (assessed by using the EWPL-TR, and the ORQ)
- b) Perceived usability (assessed by the SUS-TR)

5.2. Participants

Following table illustrates demographical information related with the survey participants. It has been observed that 50% of the participants were male, 50% of the participants were female.

Besides, 29.9% of the participants' age were between 18 and 20, 54.3% of the participants' age were between 21 and 23, 15.7% of the participants' age were over 24 (See Table 5.1).

Table 5.1. : *Demographic Information of the Working Group*

		f	%
Gender	Female	162	50
	Male	162	50
Age	18	5	1.5
	19	31	9.5
	20	61	18.8
	21	62	19.1
	22	76	23.4
	23	38	11.7
	24	23	7.0
	25	15	4.6
	26	5	1.5
	27	4	1.2
	28	3	0.9
	32	1	0.3
Total		324	100

Furthermore, 4% of the participants were in their first year , 17% of them were in their second year, 28.4% of them were in their third year, 19.8% of them were in their fourth year, 21.9% of them were in their fifth year and 9% of them were in their sixth year or more than sixth year in their education (the time they spent at the university was determined). In addition, 44.8% of the participants had majors in the departments related with computer field and 55.2% of the participants had majors in the departments that are unrelated with computer field.

When the academic grade point average (GPA) was examined for those participants, the data showed that 18.8% were below 2.00, 25.3% were between 2.00 and 2.49, 21% were between 2.50 and 2.99, 16.4% were between 3.00 and 3.49, and 7.4% were between 3.50 and 4.00, and 11.1% of the participants were not provided their GPA (See Table 5.2).

Table 5.2. : *Education Demographics of Participants*

	Items	n	%
Education Level	Undergraduate	324	100
Duration at the University	1	13	4.0
	2	55	17.0
	3	92	28.4
	4	64	19.8
	5	71	21.9
	6+	29	9.0
Total		324	100
Department	Computer field	145	44.8
	Other	179	55.2
Total		324	100
Current GPA	Below 2.00	61	18.8
	Between 2.00 and 2.49	82	25.3
	Between 2.50 and 2.99	68	21.0
	Between 3.00 and 3.49	53	16.4
	Between 3.50 and 4.00	24	7.4
	I don't know.	36	11.1

Besides, 8.6% of the participants have been using computer for 5-7 years, 21% have been using computer for 7-10 years and 70.4% have been using computer for 10 years and over. Furthermore, about the level of computer usage, it was determined that 28.4% of participants were intermediate level, 47.5% were intermediate to advanced level and 24.1% had advanced computer skills. Additionally, while participants indicated that 4.6% would generally need help in terms of computer usage, 55.2% said that it could change depending on the given task and 40.1% said they were confident in terms of computer usage.

When the participants were analyzed according to their daily average computer usage time, 54.9% of the participants stated that they use their computers 1 to 3 hours per day, 45.1% of them stated that they use their computers 4 hours and above. Based on the frequency of use of Internet-Web services by participants, it was examined that 92.6% of the participants use every day and 7.4% of the participants use a few days a week. When the frequency of using the SIS was examined, it was determined that 8.6% of the participants use the SIS every

day, 64.2% of the participants use the SIS a few days a week and 27.2% of the participants use the SIS for a few days a month. (See Table 5.3).

Table 5.3. : *Participants' Computer Related Demographic Characteristics*

	Items	n	%
Total		324	100
Web usage	Everyday	301	92.9
	Several times a week	22	6.7
	Several times a month	1	0.3
Total		324	100
Computer usage	Between three and five years	6	1.8
	Between five and seven years	22	6.7
	Between seven and Ten years	68	20.9
	Between ten years and over	228	70.4
Total		324	100
Computer skills	Intermediate level	92	28.4
	Intermediate to Advanced	154	47.5
	Advanced	78	24.1
Total		324	100
Computer usage confidence	Do not trust	1	0.3
	Usually, need help	14	4.3
	Depends on the given task	179	55.2
	Trust	130	40.1
Total		324	100
Daily computer usage	1-3 hours	178	54.9
	4 hours and over	146	45.1
Total		324	100
Internet-web services usage frequency	I use it every day.	300	92.6
	I use it a few days a week.	24	7.4
Total		324	100
SIS usage frequency	I use it every day.	28	8.6
	I use it a few days a week.	208	64.2
	I use it a few days a month.	88	27.2
Total		324	100

To analyze the details of SIS at the University, various questions have been asked to participants in detail. To illustrate, when participants asked about the most favorite features in SIS at the University, 48.1% of the participants said that none, ie. no feature is selected as a

favorite one. Besides, 23.1% those agreed with content, 14.2% those agreed with ease of use, 7.7% of those agreed with communication and 6.82% those agreed with features that are classified as other. In a similar manner, when the most disliked characteristic of SIS at the University was examined; 7.7% of the participants declared none, which means participants did not have any dislike feature at all.

Furthermore, 19.1% of those participants agreed with accessibility, 16% of those agreed with functionality, 14.9% of those agreed with login, 10.8% of those defined complexity, 6.2% of those participants accepted communication as a disliked feature, 5.9% of those agreed with content, 4.9% of those agreed with acceptance, 4.9% of those agreed with everything, 4.9% of those agreed with design, and 4.6% of those agreed with incompatibility as a disliked feature.

Finally, the necessary improvements to be done so as to boost the usability of the SIS were examined, it was determined that 37.3% of the participants mentioned development, 16% of the participants mentioned incompatibility, 10.5% of the participants selected as none means there is no need to change, 9.9% of the participants mentioned acceptance, 8.3% of the participants mentioned ease of use, 6.8% of the participants mentioned other, 4.3% of the participants mentioned content, 3.4% of the participants mentioned login, 3.4% of the participants mentioned design. After all, the most mentioned category is development and the least mentioned categories are design and login (See Table 5.4).

Table 5.4. : *Participants' Categorized Overall Reaction Questionnaire Responses*

Favorite feature in the SIS	n	%
None	156	48.1
Communication	25	7.7
Ease of use	46	14.2
Content	75	23.1
Other	22	6.8
Total	324	100.0
Most disliked features in the SIS	n	%
None	25	7.7
Communication	20	6.2
Incompatibility	15	4.6
Content	19	5.9
Functionality	52	16.0
Login	48	14.8
Accessibility	62	19.1
Everything	16	4.9
Design	16	4.9
Acceptance	16	4.9
Complexity	35	10.8
Total	324	100.0
Recommendations to improve the usability of the SIS?	n	%
None	34	10.5
Ease of use	27	8.3
Content	14	4.3
Login	11	3.4
Development	121	37.3
Design	11	3.4
Acceptance	32	9.9
Incompatibility	52	16.0
Other	22	6.8
Total	324	100.0

5.3. Equipment

In the data collection process, an electronic version of the questionnaire was prepared so that it was applied online. The address of the electronic questionnaire was sent by e-mail to the students. Electronic version of questionnaire was also shared in social media accounts. The data collection process lasted four weeks in total.

5.4. Questionnaires

Participants completed an online version of EWPL-TR, SUS-TR, ORQ and demographic surveys about their experience of the system. The questionnaires were delivered via SurveyMonkey that is an online survey package.

5.4.1. EWPL-TR

Participants' urgent emotional responses to the system were measured using a translated 15 item scale of emotion words commonly used to describe responses to sites. (For the details of the EWPL-TR see Chapter 3 and for the questions of the EWPL-TR see Table B.2. in Appendix).

5.4.2. SUS-TR

Participants' perceived usability was evaluated by the translated SUS-TR (For the details of the SUS-TR see Chapter 4 and for the questions of the SUS-TR see Table B.4. in Appendix).

5.4.3. Overall Reactions Questionnaire (ORQ)

Participants were requested to write open-ended questions about their dis/likes and suggestions on increasing usability of the SIS (for the questions of the ORQ see Table C.3 in Appendix).

5.4.4. Demographic Questionnaire

As a last step of data collection, participants completed a demographic survey that gathered information on age, sex, GPA, registered department, computer usage confidence, internet - web service usage and SIS usage (See C.4. in Appendix).

5.4.5. Procedure

The procedure for the participants was 3 steps:

Conclude the Emotional Word Prompt List (EWPL-TR) for the SIS.

Conclude the System Usability Scale (SUS-TR) for the SIS.

Conclude the Overall Reactions Questionnaire (ORQ) for the SIS.

5.5. Data Analysis

In the study, 406 undergraduate level students were participated. Subsequent to evacuating members who did not finish adequate segments of the examination, legitimate information from 324 members stayed for researching student background contrasts.

In the analysis of the data, SPSS Statistics Solution has been used and descriptive statistics such as frequency, percentage, arithmetic mean, standard deviation, minimum, maximum and parametric tests were used for those with normal distribution in the analysis of data. The two sample t-test has been used in order to simply tests whether or not two independent populations have different mean values on some measure. One-way ANOVA test was used to compare two and more independent groups, and the Scheffe test was used to compare between two independent groups. Pearson correlation coefficient was used in the analysis of the scales.

All the statistical tests in current chapter were carried out with two-tailed tests with an alpha level of 0.05 unless otherwise stated.

The next subsections will demonstrate data preparation on each variable as required.

5.5.1. Student Background

There were enough participants with valid data from different groups to create two groups of students. There were 145 participants that are educating in a major related with computer field (Information Systems and Technologies, Management Information Systems, Computer Science, Electrical Engineering, Instructional Technology and E-commerce) and 179 participants that are educating in a major unrelated with computer field.

5.5.2. Emotion Word Prompt List (EWPL-TR) Data

Appraisals on the 15 emotional words (9 positive and 6 negative) in the EWPL-TR were coded from 1 to 5. Varimax-rotation has been carried out (See glossary section for the definition of the term). According to the factor loadings which are shown in the following table, since the distances between the loads of the factors in which the items are collected must be at least 10% of the distance between them, there is no item which does not comply with this rule. As a general rule, regardless of the sign value, the load value higher than 0.60 is high, and the load value between 0.30 and 0.59 can be defined as moderate values (Buyukozturk, 2002). For SIS a clear two factors solution emerged for the questions, with all questions loading on either a positive and a negative.

Table 5.5. : *Varimax-rotated Two-Factor Result for the EWPL-TR*

Items	Factors	
	1	2
Item1	0.538	
Item6	0.582	
Item4	0.658	
Item13	0.711	
Item7	0.714	
Item14	0.720	
Item10	0.756	
Item12	0.790	
Item11	0.804	
Item2		0.726
Item3		0.783
Item5		0.652
Item8		0.782
Item9		0.816
Item16		0.668

5.5.3. System Usability Scale (SUS-TR) Data

Appraisals on the 10 words (5 positive and 5 negative) in the SUS-TR were coded from 1 to 5 (See Chapter 4 for calculation method of the SUS scores).

SUS values can range from 1 to 100 with higher values indicating higher usability (Brooke, 1996).

5.5.4. Overall Reactions Questionnaire (ORQ) Answers

Members' responses to general responses to the SIS comprised of free text. Consequently, content analysis was performed to break down this free content and an arrangement of classifications of responses was extricated.

To evaluate the precision of arrangement of members' remarks to general responses to the system, it has been ordered by author and supervisor of this proposal. Results expressed that it has been 88% agreement. That is a sufficient assention since more than 80% understanding is viewed as great.

Table 5.6. : *Category Set for Overall Reactions*

Acceptance	Comments on the adoption and use of the system, keeping it up-to-date
Accessibility	Comments on accessibility and system crash
Communication	Comments on messages, announcements and communication
Complexity	Comments on the complexity of system use
Design	Comments on system interface design
Ease of use	Comments on simple use of the system, easy use of the system
Login	Comments on the Recaptcha feature, safety questions, and system entry
Incompatibility	Comments about not opening the system on different screens, not integrating with other systems, and not having phone application
Content	Comments about the content
Development	Comments on the development of the system
Functionality	Comments on the of the system (example: course choice)
None	If (s)he says “Nothing”
Do not know	If (s)he says “I do not know”
Everything	If (s)he says “Everything”

5.6. Effects of Different Student Groups on Users' Emotion

The effects of students' demographic information on their emotion were investigated. Results on these effects will be presented in this section.

According to normal distribution analysis, it was identified that EWPL-subdimensions do not hold the conditions of normal distribution. However, other assumptions of the normal distribution were satisfied. Since kurtosis and skewness values were between ± 2.5 , mean and median values were close enough and based on central limit theorem necessity sample size was in acceptable limits (30 or more), it was determined that EWPL-TR-TR data distribution were not far away normal distribution (See Table 5.7).

Table 5.7. : *Normal Distribution Test Results of the Participants' EWPL-TR Scale*

	Statistics	Sd	P	Skewness	Kurtosis	Mean	Median
EM Positive	0.092	324	0.000	0.711	0.091	18.89	18.00
EM Negative	0.090	324	0.000	0.408	-0.499	15.09	15.00

The following table provides descriptive statistics for the EWPL-TR- scale subscales: positive and negative. To illustrate, the mean score of EM positive subscale and EM negative subscale were 18.89 ± 7.29 and 15.09 ± 6.18 , respectively (See Table 5.8).

Table 5.8. : *Descriptive Statistics of Participants' EWPL-TR Scale*

	$\bar{X}$	s.d.
EM Positive	18.89	7.29
EM Negative	15.09	6.18

According to the results of the independent sample t-test in order to test whether the participants of the survey had statistical significance of the perceptions of the EWPL subscales by gender, it was determined that there was no statistically significant difference in

the 95% confidence level between the subscales of participants' perceptions on the EWPL subscales ($t = 0.799$; $t: 0.162$; $p > 0.05$) (See Table 5.9).

Table 5.9. : *Independent Sampling T-Test Results of Differentiation of EWPL Subscales According to Gender*(* Significance Level at %95 ($P < 0.05$))

	Gender	n	$\bar{x}$	s.d.	Levene test		t	p
					F	p		
EM Positive	Female	162	18.57	7.58	1.305	.254	.799	.425
	Male	162	19.22	7.00				
EM Negative	Female	162	15.04	6.63	3.609	.058	.162	.872
	Male	162	15.15	5.72				

H₀: There were no statistically significant differences in the sub-dimension perceptions of the participants on the EWPL scale according to gender.

H₁: There were statistically significant differences in the sub-dimension perceptions of the participants on the EWPL scale according to gender.

According to the results of the one-way ANOVA test in order to determine whether the perceptions of the subscales of the EWPL scale were statistically significant by age groups, the subscales of the EWPL scale were not statistically significant at 95% confidence level ($p > 0.05$) (See Table 5.10).

Table 5.10. : *Results of One-Way ANOVA Test for Differentiation of EWPL Subscales According to Age Groups*(* Significance Level at %95 ($P < 0.05$))

	Age	n	$\bar{x}$	s.d.	F	p
EM Positive	18-20	97	20.13	7.68	2.021	0.134
	21-23	176	18.35	7.02		
	24 and over	51	18.41	7.32		
EM Negative	18-20	97	15.68	5.78	1.153	0.317
	21-23	176	15.08	6.29		
	24 and over	51	14.06	6.53		

H₀:The subscale scores of participants' EM scale were not statistically significant according to age groups.

H₁:The subscale scores of participants' EM scale were statistically significant according to age groups.

The one-way ANOVA test results presented in the following table, determines whether the survey participants' perceptions of the dimensions of the EWPL-TR scale were statistically significant or not based on participant' experience (in terms of year) in the university. According to the obtained results, the perceptions belonging to EM positive sub-dimension were 95% statistically significant compared to the various groups under distinct values of duration variable at the University ($F = 2.951$, $p < 0.05$).

The variance was found to be homogeneous ($p = 0.562$; $p > 0.05$) as a result of the homogeneity test of the variance of the participants' EM positive perceptions. For this purpose, the Scheffe test, which is one of the post hoc tests, was considered to determine the source of the variance.

Based on the Scheffe test results, there exists a significant difference between the students who are in their fifth year and first, second and third years. Besides, the EM positive perceptions of the fifth year students ($=16.42$) are lower than those of the first year ($=21.76$), second year ($=20.65$) and third year ($=19.58$) students (See Table 5.11).

Table 5.11. : One-Way ANOVA Test Results for EWPL-TR Scale measured Under Different Duration Sizes at the University (* Significance Level at %95 (P<0.05))

	Duration at University	n	$\bar{x}$	s.d.	F	p	Scheffe
EM Positive	1 ⁽¹⁾	13	21.76	8.23	2.951	0.013 *	(1-5) (2-5) (3-5)
	2 ⁽²⁾	55	20.65	7.47			
	3 ⁽³⁾	92	19.58	7.10			
	4 ⁽⁴⁾	64	18.68	7.44			
	5 ⁽⁵⁾	71	16.42	6.98			
	6+ ⁽⁶⁾	29	18.55	6.24			
EM Negative	1	13	17.92	4.64	1.488	0.193	-
	2	55	15.78	5.81			
	3	92	15.47	6.20			
	4	64	15.20	6.23			
	5	71	13.80	6.39			
	6+	29	14.27	6.44			

Based on different durations at the University, the following hypothesis test has been conducted for both EM Positive and Negative EWPL-TR Scales.

H₀: The differences of participants' EWPL-TR perceptions based on participant' experience (in terms of year) in the university are not statistically significant.

H₁: The differences of participants' opinions on the EWPL-TR scale based on participant' experience (in terms of year) in the university are statistically significant.

According to the results of the independent sample t-test presented in the following table., for testing the statistical significance of the differences of subscales of EWPL-TR scale of participants in the survey, the difference of EWPL-TR positive subscale perceptions of participants was determined as statistically significant at 95% confidence level ($t = 2.026$; $p < 0.05$). Besides, those who are registered in a major that is related with computer field ($\bar{x} = 19.80$) have higher EM positive perceptions than those registered in a major that is unrelated with computer field ($\bar{x} = 18.15$). Similarly, the difference of the EM negative subscale perceptions with respect to the registered department was statistically significant at the 95%

confidence level ($t = 2.231$; $p < 0.05$). Additionally, the EM negative perceptions of those registered in a major that is related with computer field ($\bar{x}=15.94$) are higher than those registered in a major that is unrelated with computer field ($\bar{x}=14.41$) (See Table 5.12).

Table 5.12. : *Independent Sampling T-Test Results for Status of the EWPL-TR Scale Under different Registered Departments (* Significance Level at %95 ($P < 0.05$))*

	Registered Departmen t	n	$\bar{x}$	s.d.	Levene Test		t	P
					F	p		
EM Positive	Computer Field	14 5	19.80	7.72	1.281	0.258	2.026	0.044*
	Other	17 9	18.15	6.86				
EM Negative	Computer Field	14 5	15.94	6.66	3.865	0.050	2.231	0.026*
	Other	17 9	14.41	5.68				

H₀: The differences of the participants' perceptions of the EWPL-TR scale compared to the registered department are not statistically significant.

H₁: The differences of the participants' perceptions of the EWPL-TR scale compared to the registered department are statistically significant.

According to the results of the one-way ANOVA test, in the following table, EM positive dimension perceptions were found to be statistically significant at 95% confidence level compared to the GPA groups ($F = 3.487$; $p < 0.05$). In addition to this result, the variance of participants' EM positive perceptions' was found to be homogeneous ($p = 0.083$; $p > 0.05$) as a result of the Scheffe homogeneity test. In the last column of the table., Scheffe test results showed that there is a significant difference between the GPA groups that are "I do not know" and the other groups that are "Below 2.00", "Between 2.00 and 2.49", "Between 2.50 and 2.99", "Between 3.00 and 3.49", "Between 3.50 and 4.00". EM positive perceptions of students who do not know their GPA ($\bar{x}=23.02$) are higher than those who have GPA ranged

from 2.00 to 2.49 ($\bar{x}$ =19.63) , from 2.50 to 2.99 ($\bar{x}$ =18.26) , from 3.00 to 3.49 ($\bar{x}$ =17.52) , 3.50 and 4.00 ($\bar{x}$ =17.95) (See Table 5.13).

Table 5.13. : One-way ANOVA test results for differences of EWPL-TR Sub-Scales according to the overall GPA (* Significance Level at %95 (P<0.05))

	Grade Point Average	n	$\bar{x}$	s.d.	F	p	Scheffe
EM Positive	Below 2.00 (1)	61	17.70	6.22	3.487	0.004 *	(1-6) (2-6) (3-6) (4-6) (5-6)
	Between 2.00 and 2.49 (2)	82	19.63	7.86			
	Between 2.50 and 2.99 (3)	68	18.26	6.69			
	Between 3.00 and 3.49 (4)	53	17.52	6.93			
	Between 3.50 and 4.00 (5)	24	17.95	6.57			
	Don't know (6)	36	23.02	8.37			
EM Negative	Below 2.00	61	15.22	5.83	2.121	0.063	-
	Between 2.00 and 2.49	82	15.86	6.72			
	Between 2.50 and 2.99	68	14.13	6.07			
	Between 3.00 and 3.50	53	13.60	5.99			
	Between 3.50 and 4.00	24	14.95	6.43			
	Don't know	36	17.25	5.18			

To understand the impact of overall Grade Point Average (GPA) over the perceptions of participants, the following hypothesis has been tested at %95 significance level, similar to previous analysis.

H₀: The differences of participants' EWPL-TR perceptions according to the GPA are not statistically significant.

H₁: The differences of participants' EWPL-TR perceptions according to the GPA are statistically significant.

According to result of One-Way ANOVA test, presented in the following table, hypothesis test was considered to understand the effect of the computer usage time on the participants' perceptions in both subscales of EWPL-TR. Based on the tested hypothesis and the corresponding one-way ANOVA test results, it was determined that the EM positive dimension perceptions were statistically significant at 95% ($F= 5.530$; $p<0.05$) in the following table. In addition to this result, with the help of Scheffe homogeneity test, the variance of the participants' EM positive was found to be homogeneous ($p = 0.936$; $p> 0.05$). As an another conclusion based on the Scheffe test, there is a significant difference between the participants who have 5-7 years computer usage experience and those who have experience between 7-10 years, 10 and more than that. The EM positive perceptions of students who have 5-7 years experience on computer usage ($=23.00$) are higher than those who use computers between 7-10 years ($=19.29$), 10 years and over 10 years ($=18.26$) (See Table 5.14).

Table 5.14. : *One-Way ANOVA Test Results for Differentiation of Dimensions of EWPL-TR Scale according to Computer Usage Time (* Significance Level at %95 (P<0.05))*

	Computer Usage Time	n	$\bar{x}$	s.d.	F	p	Scheffe
EM Positive	I use it for 5-7 years ⁽¹⁾	28	23.00	7.88	5.530	0.004 *	(1-2) (1-3)
	I use it for 7-10 years ⁽²⁾	68	19.29	6.86			
	I use it for 10 years and over. ⁽³⁾	228	18.26	7.20			
EM Negative	I use it for 5-7 years	28	17.00	6.36	1.755	0.175	-
	I use it for 7-10 years	68	15.42	6.14			
	I use it for 10 years and over.	228	14.76	6.14			

H₀: Participants' EWPL-TR perceptions do not differ significantly by the computer usage time.

H₁: Participants' EWPL-TR perceptions differ significantly by the computer usage time.

According to the results of the one-way ANOVA test, presented in the following table, the EWPL-TR positive dimension perceptions were found to be statistically significant at 95% confidence level according to the group ($F= 3.043$; $p< 0.05$). As a result of the Scheffe homogeneity test, the variance of the participants' perceptions of EM positive was found to be homogeneous ($p = 0.225$; $p> 0.05$). Besides, again by using the Scheffe test, it is showed that there is a significant difference on the perceptions of the participants who have either advanced and intermediate-advanced levels of computer skills. To sum up, the EM positive perceptions of students who have advanced computer usage skills ($=17.12$) are lower than those of have intermediate level ($=19.41$) and intermediate-advanced ($=19.47$) computer usage skills (See Table 5.15).

Similar to the result of EM positive, EM negative perceptions were found to be statistically significant at 95% confidence level according to various computer usage skills ($F=6.379$; $p<0.05$). The variance of the participants' EM negative was found to be homogeneous as a result of the Scheffe homogeneity test ($p = 0.106$; $p> 0.05$). Additionally, there is a significant difference between advanced level and levels that are intermediate and intermediate-advanced levels of computer usage skills. The EM negative perceptions of students who have advanced computer usage skills ($\bar{x}=13.21$) are lower than those of intermediate level ($\bar{x}=16.55$) and intermediate-advanced ($\bar{x}=15.18$) (See Table 5.15).

Table 5.15. : *One-Way ANOVA Test Results for Differentiation of the EWPL-TR Sub-Scales with respect to the Computer Usage Skills (* Significance Level at %95 ($P<0.05$))*

	Computer Usage Skills	n	$\bar{x}$	s.s.	F	p	Scheffe
EM Positive	Intermediate ⁽¹⁾	92	19.41	6.98	3.043	0.049 *	(1-3) (2-3)
	Intermediate - Advanced ⁽²⁾	154	19.47	7.68			
	Advanced ⁽³⁾	78	17.12	6.62			
EM Negative	Intermediate ⁽¹⁾	92	16.55	6.20	6.379	0.002 *	(1-3) (2-3)
	Intermediate - Advanced ⁽²⁾	154	15.18	6.45			
	Advanced ⁽³⁾	78	13.21	5.08			

In a similar manner, the impact of the computer usage skills were tested with the following hypothesis test.

H₀: The differences of EWPL-TR perceptions of participants according to their computer skills were not statistically significant.

H₁: Participants' EWPL-TR perceptions differ statistically according to their computer skills.

Based on the under mentioned hypothesis test, One-way ANOVA test was conducted in order to test whether there exists a significant difference among the perceptions of EWPL-TR sub-scales based on self-confidence on computer usage. According to the presented results in Table 5.13., the EM negative dimension perceptions were statistically significant, 95% confidence level ($F = 3.188$, $p < 0.05$).

Similar to the previous calculations, the variance of the participants' who have EM negative perceptions were found to be homogeneous as a result of Scheffe homogeneity test ($p = 0.422$; $p > 0.05$). Furthermore, as a result of the Scheffe test, there is a significant difference between the participants who declared “I usually need help” and “I trust” phrases about their computer usage. The mean values for those who said “I usually need help” ($=18.26$) of EM Negative perceptions are higher than those who said “I trust” ($=14.33$) about their computer usage (See Table 5.16).

Table 5.16. : *One-way ANOVA Test Results for the Differentiation of Dimensions of EWPL-TR Scale According to Self-Confidence Variation on Computer Usage (* Significance Level at %95 ($P < 0.05$))*

	Self-confidence variation on computer usage	n	$\bar{x}$	s.d.	F	p	Scheffe
EM Positive	I usually need help	15	19.73	5.25	1.347	0.262	-
	It depends on the given task.	179	19.40	7.55			
	I trust.	130	18.08	7.09			
EM Negative	I usually need help. ⁽¹⁾	15	18.26	5.45	3.188	0.043 *	(1-3)
	It depends on the given task. ⁽²⁾	179	15.38	6.23			
	I trust ⁽³⁾	130	14.33	6.08			

In this step, the effect of self-confidence variation on computer usage for both EM positive and negative were tested within the following hypothesis test,

H₀: There is no statistically significant difference between participants' perceptions of EWPL-TR sub-scales based on self-confidence on computer usage.

H₁: There is a significant difference among the participants' perceptions of EWPL-TR sub-scales with respect to the self-confidence on computer usage.

The following table presents the results of the independent sample t-test presented that the difference of the EM positive subscale perceptions according to the internet web service usage frequency was statistically significant at 95% confidence level ($t = -2.331$, $p < 0.05$). Besides, participants who use internet web services a few days in a week ($\bar{x}=22.20$) have higher EM positive perceptions than those who use such services daily ($\bar{x}=18.62$) (See Table 5.17).

Table 5.17. : *Independent Sampling T-Test Results For The Differentiation of EWPL-TR Scale Based On The Internet Web Service Usage Frequency (* Significance Level at %95 (P<0.05))*

	Internet-Web Service Usage Frequency	n	$\bar{x}$	s.d.	Levene Test		t	P
					F	p		
EM Positive	Everyday	300	18.62	7.25	0.006	0.936	-2.331	0.020*
	Several days a week	24	22.20	7.07				
EM Negative	Everyday	300	14.92	6.15	0.171	0.679	-1.847	0.066
	Several days a week	24	17.33	6.17				

H₀: Participants' EWPL-TR perceptions are not significantly different according to the internet web service usage frequency.

H₁: Differences in participants' EWPL-TR perceptions are statistically significant according to the internet web service usage frequency.

To determine whether the survey participants' perceptions of the EWPL-TR sub-scales were statistically significant or not, one-way ANOVA test was considered and its results were presented in the following table. Based on the obtained test results, it was determined that EM positive dimension perceptions were statistically significant at the 95% confidence level according to the frequency of SIS use at the University ($F = 6.575$; $p < 0.05$).

Afterward, the Tamhane test was used as the post hoc test to determine the source of the variance because of the nonhomogeneous variance of the participants' ($p = 0.002$; $p < 0.05$). As a result of the this test, there is a significant difference between participants those who use SIS a few days in a month and those who use it every day and a few days a week. Besides, EM positive perceptions of students using a few days per month ($\bar{x}=16.54$) are lower than those using a few days per week ($\bar{x}=20.35$), and a few days per week ($\bar{x}=19.68$).

Similarly, EM negative perceptions were found to be statistically significant at 95% confidence level ($F = 3.714$; $p < 0.05$) for SIS usage frequency groups at the University.

Participants' variance for the EM negative perceptions was analyzed by using the Scheffe homogeneity test to determine the source of variance when the variances were homogeneous ($p = 0.719$; $p > 0.05$). As a result of the Scheffe test, there exists a significant difference between the participants those who use SIS a few days a week and a few days a month. To sum up, EM negative perceptions of students who use a few days per month ($\bar{x}=13.60$) are lower than those who use a few days per week ($\bar{x}=15.58$), as shown in the following table (See Table 5.18).

Table 5.18. : One-way ANOVA test results for differentiation of dimensions of the EWPL-TR scale based on the University' SIS Usage Frequency (* Significance Level at %95 (P<0.05))

	University' SIS Usage Frequency	n	$\bar{x}$	s.d.	F	p	Scheffe
EM Positive	I use it everyday. (1)	28	20.35	8.77	6.575	0.002 *	(1-3) (2-3)
	I use it a few days a week. (2)	208	19.68	7.41			
	I use it a few days a month. (3)	88	16.54	5.91			
EM Negative	I use it everyday. (1)	28	16.17	7.08	3.714	0.025 *	(2-3)
	I use it a few days a week. (2)	208	15.58	6.10			
	I use it a few days a month. (3)	88	13.60	5.85			

Specifically, the impact of the usage of SIS at the University over the perceptions of the participants has been tested via the below-described hypothesis test,

H₀: There is no statistically significant difference between perceptions of participants' EM sub-scale according to SIS usage frequency at the University.

H₁: There exists a statistically significant difference between perceptions of participants' EM sub-scale according to SIS usage frequency at the University.

According to the results of the one-way ANOVA test, presented in the following table, the perceptions of participants of the EWPL-TR subscales were compared and it was determined that the most liked features of EM positive perceptions of the University SIS statistically significant at 95% confidence level ($F = 10.130$, $p < 0.05$). Because of the homogeneity test for the variance of the participants' EM positive perceptions, the Scheffe test was used to determine the source of variance when the variances were homogeneous ($p = 0.514$; $p > 0.05$). Once the Scheffe test was considered, the obtained results expressed a

significant difference between features of SIS like none, communication, ease of use, content, and other, on the participant perceptions. The mean value of EM Positive perceptions of having no liked feature (=**16.44**) are lower than the participant perceptions who have most liked features are communication (=**20.32**), ease of use (=**22.17**), content (=**20.48**), and other (=**22.36**).

EM negative dimension perceptions were statistically significant at 95% confidence level ($F = 8.928$, $p < 0.05$), compared to the most liked features groups at the University SIS. Similar to EM positive case, the Scheffe test of post hoc tests was used to determine the source of variance ($p = 0.579$; $p > 0.05$) as a result of the homogeneity test of the variance of EM negative perceptions. The result of Scheffe test showed that there exists a significant difference between features of SIS like none, communication, ease of use, content, and other, on the participant perceptions. The mean value of EM Negative perceptions of having no liked feature (=**13.30**) are lower than the participant perceptions who have most liked features are ease of use (=**18.23**), content (=**16.17**), other (=**18.09**). Additionally, EM Negative perceptions of the participants who have most liked feature is communication (=**14.64**) are lower than those who have ease of use (=**18.23**) and other features (=**18.09**) (See Table 5.19).

Table 5.19. : *One-way ANOVA Test Results for Differentiation of EWPL-TR Sub-Scales according to the Most Liked Features of SIS at the University (* Significance Level at %95 (P<0.05))*

	University'SIS Most Liked Features	n	$\bar{x}$	s.d.	F	p	Scheffe
EM Positive	None ⁽¹⁾	156	16.44	6.38	10.13 0	0.000 *	
	Communication ⁽²⁾	25	20.32	7.35			(1-2)
	Ease of use ⁽³⁾	46	22.17	7.83			(1-3)
	Content ⁽⁴⁾	75	20.48	7.07			(1-4)
	Other ⁽⁵⁾	22	22.36	7.44			(1-5)
EM Negative	None ⁽¹⁾	156	13.30	5.48	8.928	0.000 *	
	Communication ⁽²⁾	25	14.64	6.10			(1-3)
	Ease of use ⁽³⁾	46	18.23	6.75			(1-4)
	Content ⁽⁴⁾	75	16.17	5.89			(1-5)
	Other ⁽⁵⁾	22	18.09	6.62			(2-3)
							(2-5)

The impact of the favorite features of SIS at the University on the participant perceptions has been conducted with the following hypothesis test.

H₀: There is no statistically significant difference between participants' perceptions of the EWPL-TR subscale according to the most liked features of the University SIS.

H₁: Participants' perceptions of the EWPL-TR sub-scales differ significantly according to the most liked features of the University SIS.

The following table presents the results of the one-way ANOVA test in order to test the statistical significance of the difference according to the most disliked features of the University'SIS were summarized. Based on the comparison of participants' perceptions of EWPL-TR subscales according to the most disliked features of SIS at the University, there exists a statistically significant difference at the 95% confidence level ($F = 3.525$, $p < 0.05$). Similarly, the Scheffe test was used to determine the source of the variance ($p = 0.076$; $p > 0.05$) as the variance homogeneity test for the variance of the participants' EM positive

perceptions was homogeneous. This test results in a significant difference between none and communication, incompatibility, content, functionality, login, accessibility, everything, design, acceptance, with respect to their impacts on the participants' perceptions of EWPL-TR subscales.

The mean value of EM Positive perceptions of having no disliked feature (**=24.00**) were higher than others who have disliked features such as communication (**=18.20**), incompatibility (**=18.40**), content(**=17.31**), functionality (**=18.96**), login (**=17.95**), accessibility (**=19.54**), everything (**=12.50**), design (**=19.43**), acceptance (**=15.68**). Another result showed that, the mean value of EM Positive perceptions of having disliked feature as "Everything" (**=12.50**) were lower than others who have disliked features communication (**=18.20**), incompatibility (**=18.40**), content (**=17.31**), functionality (**=18.96**), login (**=17.95**), accessibility (**=19.54**), design (**=19.43**), acceptance (**=15.68**), complexity (**=20.85**).

When EM negative perceptions were considered, a similar result was found that there is a statistically significant difference at 95% confidence level ($F = 3.465$; $p < 0.05$), according to the most disliked features of the SIS at the University. Similar to the result of EM Positive, the Scheffe test of post hoc tests was used to determine the source of variance since the homogeneity test of variance of EM negative perceptions of participants was found to be homogeneous ($p = 0.109$; $p > 0.05$). As a result of the Scheffe test, there is a significant difference between disliked features like none and incompatibility, content, functionality, login, accessibility, everything, acceptance, complexity.

To illustrate, the mean value of EM Negative perceptions of having no disliked feature (**=19.76**) were higher than those who have incompatibility (**=14.26**), content (**=14.63**), functionality (**=14.61**), login (**=14.81**), accessibility (**=14.25**), everything (**=9.50**), acceptance (**=15.48**), complexity (**=15.94**) disliked features for SIS. Another result shows that, the mean

value of EM Negative perceptions of having disliked feature as “Everything” (=9.50) were lower than those who have disliked features such as communication (=16.90), incompatibility (=14.26), content (=14.3), functionality (=14.61), login (=14.81), accessibility (=14.25), design (=16.25), acceptance (=15.18), complexity (=15.94) (See Table 5.20).

Table 5.20. : One-way ANOVA Test Results for Differentiation of EWPL-TR Subscales under the Most Disliked Features of SIS at the University (* Significance Level at %95 (P < 0.05))

	Most Disliked Features in the University'SIS	n	$\bar{x}$	s.d.	F	p	Scheffe
EM Positive	None ⁽¹⁾	25	24.00	8.72	3.525	0.000 *	(1-2)
	Communication ⁽²⁾	20	18.20	6.58			(1-3)
	Incompatibility ⁽³⁾	15	18.40	8.36			(1-4)
	Content ⁽⁴⁾	19	17.31	7.21			(1-5)
	Functionality ⁽⁵⁾	52	18.96	6.77			(1-6)
	Login ⁽⁶⁾	48	17.95	7.79			(1-7)
	Accessibility ⁽⁷⁾	62	19.54	7.22			(1-8)
	Everything ⁽⁸⁾	16	12.50	4.16			(1-9)
	Design ⁽⁹⁾	16	19.43	6.12			(1-10)
	Acceptance ⁽¹⁰⁾	16	15.68	5.09			(2-8)
	Complexity ⁽¹¹⁾	35	20.85	6.15			(3-8)
EM Negative	None ⁽¹⁾	25	19.76	6.79	3.465	0.000 *	(4-8)
	Communication ⁽²⁾	20	16.90	6.78			(5-8)
	Incompatibility ⁽³⁾	15	14.26	6.21			(6-8)
	Content ⁽⁴⁾	19	14.63	6.30			(7-8)
	Functionality ⁽⁵⁾	52	14.61	6.56			(8-8)
	Login ⁽⁶⁾	48	14.81	6.20			(9-8)
	Accessibility ⁽⁷⁾	62	14.25	5.64			(10-8)
	Everything ⁽⁸⁾	16	9.50	4.38			(11-8)
	Design ⁽⁹⁾	16	16.25	5.89			
	Acceptance ⁽¹⁰⁾	16	15.18	4.81			
	Complexity ⁽¹¹⁾	35	15.94	4.75			

Again for the SIS at the University, how the participants' perceptions are affected by the most disliked features was evaluated by using the following test,

H₀: The difference of participants' EWPL-TR perceptions according to the most disliked features of SIS is not statistically significant.

H₁: The differences of participants' EWPL-TR perceptions according to the most disliked feature of SIS are statistically significant.

In order to test whether there exists a statistically significant difference among the perceptions of EWPL-TR subscales according to the recommendations to increase the usability of SIS at the University, the one-way ANOVA test was conducted and its results were given in the following table(See Table 5.21). According to the obtained results of one-way ANOVA, EM positive perceptions were statistically significant 95% confidence level ($F = 2.779$, $p < 0.05$). As a result of the homogeneity test for checking the variance of the participants' EWPL-TR positive, it was found to be homogeneous ($p = 0.159$; $p > 0.05$) and for this reason again Scheffe test was used to determine the source of the variance. The Scheffe test results showed that there exists a significant difference between the participants' perceptions based on the recommendations like development, incompatibility, content with development, incompatibility, design with development, incompatibility with none.

The mean value of EM Positive perceptions with respect to none of the recommendations (**=21.82**) were higher than the other suggestions such as development (**=18.16**), incompatibility (**=16.28**) for the increase usage of SIS. Besides, EM Positive perceptions of the participants who prefer content (**=22.92**) were higher than those who prefer development (**=18.16**) and incompatibility (**=16.28**). Furthermore, EM Positive perceptions of

the participants who prefer design (**=22.72**) were higher than those who prefer development (**=18.16**), incompatibility (**=16.28**).

Additionally, EM negative perceptions were found to be statistically significant at 95% confidence level ($F = 2.571$; $p < 0.05$) based on the recommendations in order to increase the usability of SIS at the University. The variance of the participants' EM negative perceptions was found to be homogeneous ($p = 0.151$; $p > 0.05$) and the Scheffe test was used to determine the source of variance. As a result, there is a significant difference between the participants' perceptions based on the recommendations like development and none, content, design, and other. For instance, the mean value of EM Negative perceptions with respect to Development (**=13.58**) were lower than those prefer none (**=17.20**), content (**=18.07**), design (**=17.81**) other (**=17.40**).

Table 5.21. : One-way ANOVA Test Results of Differentiation of the EWPL-TR Subscales according to Recommendations about Increase Usage of SIS at the University (* Significance Level at %95 ($P < 0.05$))

	Recommendations for Increasing the Usability of the University'SIS	n	$\bar{x}$	s.d.	F	p	Scheffe
EM Positive	None ⁽¹⁾	34	21.82	8.15	2.779	0.006 *	
	Ease of use ⁽²⁾	27	19.11	7.36			
	Content ⁽³⁾	14	22.92	9.40			(1-5)
	Login ⁽⁴⁾	11	20.00	8.02			(1-8)
	Development ⁽⁵⁾	121	18.16	7.08			(3-5)
	Design ⁽⁶⁾	11	22.72	8.68			(3-8)
	Acceptance ⁽⁷⁾	32	18.53	6.26			(5-6)
	Incompatibility ⁽⁸⁾	52	16.28	6.11			(6-8)
	Other ⁽⁹⁾	22	19.72	5.99			
EM Negative	None ⁽¹⁾	34	17.20	6.79	2.571	0.010 *	
	Ease of use ⁽²⁾	27	15.03	4.38			(1-5)
	Content ⁽³⁾	14	18.07	7.77			(3-5)
	Login ⁽⁴⁾	11	15.09	6.04			(6-5)
	Development ⁽⁵⁾	121	13.58	6.12			(9-5)

Design ⁽⁶⁾	11	17.81	4.66
Acceptance ⁽⁷⁾	32	15.31	6.07
Incompatibility ⁽⁸⁾	52	14.78	5.72
Other ⁽⁹⁾	22	17.40	6.48

The impact of above-tabulated recommendations on the increasing usage of SIS was tested with the following hypothesis test,

H₀: Participants' perceptions of EWPL-TR subscales are not statistically significant with respect to the recommendations to increase the usage of SIS at the University.

H₁: Participants' perceptions of EWPL-TR subscale differ statistically with respect to the recommendations to increase the usage of SIS at the University.

5.7. Effects of Different Student Groups on Users' Perceived Usability

The effects of different student groups' background on users' perceived usability were investigated. Results will be presented in this section.

According to normal distribution analysis, it was identified that SUS-TR scale hold the conditions of normal distribution ($p: 0.070$, $p > 0.05$, See Table 5.22).

Table 5.22. : *Normal Distribution Test Results of the Participants' SUS-TR Scale*

	Statistics	n	p	Skewness	Kurtosis	Mean	Median
SUS-TR	0,051	324	0,070	-0,081	-0,261	46,30	47,50

The following table provides descriptive statistics for the SUS-TR scale. To illustrate, the mean score of SUS-TR were 46.30 ± 20.78 (See Table 5.20). Participants' mean SUS-TR scores belongs to "Not Acceptable" class which is between "Poor" and "OK".

Table 5.23. : *Descriptive Statistics of Participants' SUS-TR Scale*

	$\bar{X}$	s.d.
SUS-TR	46,30	20,78

According to the independent sample T-Test to test whether the SUS scores of the participants in the survey based on the gender were statistically significant, it was determined that the SUS-TR scores of the participants based on the gender were not statistically significant at the 95% confidence level ($t = 0.327$; $p > 0.05$), (See Table 5.24).

Table 5.24. : *Independent Sampling T-Test Results for Differentiation Status of the SUS Scale According to the Gender(* Significance Level at %95 ($P < 0.05$))*

	Gender	n	$\bar{x}$	s.d.	Levene Test		t	P
					F	p		
SUS-TR	Female	162	45.93	22.36	1.903	0.169	0.327	0.744
	Male	162	46.68	19.15				

H₀: There was no statistically significant difference in the total score of participants' SUS-TR scale by gender.

H₁: The differences in the total score of participants' SUS –TR scale by gender were statistically significant.

According to the results of the one-way ANOVA test in order to test whether the differences of the SUS-TR scores of the participants according to age groups were statistically significant , it was determined that the scores of the SUS-TR scale were not statistically significant at the 95% confidence level according to age groups ($F = 2.238$, $p > 0.05$) (See Table 5.25).

Table 5.25. : *One-way ANOVA test results for differentiation of SUS-TR scale according to age groups*(* Significance Level at %95 (P<0.05))

	Age	n	$\bar{x}$	s.d.	F	p
SUS-TR	18-20	97	57.31	9.06	2.238	0.108
	21-23	176	59.66	8.70		
	24+	51	57.70	12.13		

H₀: The total score of the participants on the SUS-TR scale was not statistically significant according to age groups.

H₁: The total score of the participants on the SUS-TR scale was statistically significant according to age groups.

The one-way ANOVA test results presented in the following table, determines whether the survey participants' scores of the SUS-TR scale were statistically significant or not based on participant' experience (in terms of year) in the university. According to the obtained results, the scores belonging to SUS-TR scale were not statistically significant 95% compared to the various groups under distinct values of duration variable at the University (F=1.137; p > 0.05) (See Table 5.26).

Table 5.26. : *One-Way ANOVA Test Results for SUS-TR Scale measured Under Different Duration Sizes at the University* (* Significance Level at %95 (P<0.05))

	Duration at the University	n	$\bar{x}$	s.d.	F	p
SUS-TR	1	13	56.34	10.19	1.137	0.340
	2	55	57.23	9.07		
	3	92	58.02	7.83		
	4	64	59.06	9.56		
	5	71	60.63	9.88		
	6+	29	58.62	12.71		

H₀: The differences of participants' SUS-TR scores based on participant' experience (in terms of year) in the university are not statistically significant.

H₁: The differences of participants' scores on the SUS-TR scale based on participant' experience (in terms of year) in the university are statistically significant.

According to the results of the independent sample t-test presented in the following table, for testing the statistical significance of the SUS-TR scale scores of participants in the survey, the SUS scores were not statistically significant at 95% confidence level ($F = 0.851$; $p > 0.05$) (See Table 5.27).

Table 5.27. : *Independent Sampling T-Test Results for the SUS-TR Scale Under different Registered Departments (* Significance Level at %95 ($P < 0.05$))*

	Registered Department	N	$\bar{x}$	s.d.	Levene test		t	p
					F	p		
SUS-TR	IT	145	47.40	20.77	0.081	0.776	0.851	0.395
	Others	179	45.42	20.81				

H₀: The differences of the participants' scores of the SUS-TR scale compared to the registered department are not statistically significant.

H₁: The differences of the participants' scores of the SUS-TR scale compared to the registered department are statistically significant.

According to the results of the one-way ANOVA test , in the following table, in order to determine whether the scores of the survey participants were statistically significant according to the overall GPA, it was determined that the SUS scale scores were not statistically significant at 95% confidence level according to the groups according to the overall GPA ($F = 0.238$; $p > 0.05$) (See Table 5.28).

Table 5.28. : *One-way ANOVA test results for differences of SUS-TR Scale according to the overall GPA(* Significance Level at %95 (P<0.05))*

	GPA	n	x	s.d.	F	p
SUS-TR	Below 2.00	61	46.48	21.79	0.238	0.945
	Between 2.00 and 2.49	82	46.28	21.53		
	Between 2.50 and 2.99	68	47.21	19.52		
	Between 3.00 and 3.49	53	47.74	22.84		
	Between 3.50 and 4.00	24	44.06	19.64		
	I do not know.	36	43.75	18.08		

H₀: The differences of participants' SUS-TR scores according to the GPA are not statistically significant.

H₁: The differences of participants' SUS-TR scores according to the GPA are statistically significant.

According to the results of the one-way ANOVA test in order to test the statistical significance of the SUS scores of the participants according to the their computer usage time, the SUS scores of the participants were not statistically significant at 95% confidence level ($F = 2.342$; $p > 0.05$) (See Table 5.29).

Table 5.29. : *One-Way ANOVA Test Results for SUS-TR Scale according to Computer Usage Time (* Significance Level at %95 (P<0.05))*

	Computer Usage Time	n	$\bar{x}$	s.d.	F	p
SUS-TR	I use it for 5-7 years.	28	45.63	23.66	2.342	0.098
	I use it for 7-10 years.	68	41.62	19.13		
	I use it for 10 years and over.	228	47.79	20.77		

H₀: Participants' SUS-TR perceptions do not differ significantly by the computer usage time.

H₁: Participants' SUS-TR perceptions differ significantly by the computer usage time.

According to the results of the one-way ANOVA test in order to test whether statistical significance of the SUS-TR scores of the participants in the survey according to the computer usage skills was significant, it was determined that the scores of the SUS-TR scale were not statistically significant at the 95% confidence level according to computer usage skills ($F = 1.230$; $p > 0.05$) (See Table 5.30).

Table 5.30. : *One-Way ANOVA Test Results for Differentiation of the SUS-TR Scale with respect to the Computer Usage Skills (* Significance Level at %95 ($P < 0.05$))*

	Computer Usage Skills	n	$\bar{x}$	s.d.	F	p	Difference
SUS-TR	Intermediate	92	60.08	8.63	1.230	0.230	-
	Intermediate - Advanced	154	58.07	9.18			
	Advanced	78	10.79	4.52			

* $p < 0.05$

In a similar manner, the impact of the computer usage skills were tested with the following hypothesis test.

H₀: The differences of SUS-TR perceptions of participants according to their computer skills were not statistically significant.

H₁: Participants' SUS-TR perceptions differ statistically according to their computer skills.

According to the results of the one-way ANOVA test in order to test the statistical significance of the SUS scores of participants in the survey according to their self-confidence in computer usage, it was determined that the scores of the SUS scale were not statistically significant at the 95% confidence level ($F=2.359$; $p>0.05$) (See Table 5.31).

Table 5.31. : *One-way ANOVA Test Results for the SUS-TR Scale According to Self-Confidence Variation on Computer Usage (* Significance Level at %95 (P<0.05))*

Self-confidence variation on computer usage		n	$\bar{x}$	S.d.	F	p
SUS-TR	I usually need help.	15	56.83	23.21	2.359	0.096
	It depends on the given task.	179	44.99	20.64		
	I trust.	130	46.90	20.48		

H₀: There is no statistically significant difference between participants' scores of SUS-TR scale based on self-confidence on computer usage.

H₁: There is a significant difference among the participants' scores of SUS-TR scale with respect to the self-confidence on computer usage.

The following table presents the results of the independent sample t-test presented that the difference of the SUS-TR scores according to the internet web service usage frequency was not statistically significant at 95% confidence level ($t=0.166$; $p > 0.05$) (See Table 5.32).

Table 5.32. : *Independent Sampling T-Test Results For The Differentiation of SUS-TR Scale Based On The Internet Web Service Usage Frequency (* Significance Level at %95 (P<0.05))*

	Internet- Web Service Usage Frequency	n	$\bar{x}$	s.d.	Levene Testi		t	p
					F	p		
SUS-TR	Everyday	300	46.36	20.91	0.847	0.358	0.166	0.868
	Several times a week	24	45.63	19.51				

H₀: Participants' SUS-TR scores are not significantly different according to the internet web service usage frequency.

H₁: Differences in participants' SUS-TR scores are statistically significant according to the internet web service usage frequency.

To determine whether the survey participants' scores of the SUS-TR scales were statistically significant or not, one-way ANOVA test was considered and its results were presented in the following table. Based on the obtained test results, it was determined that SUS-TR scores were not statistically significant at the 95% confidence level according to the frequency of SIS use at the University ($F=1.359$; $p > 0.05$) (See Table 5.33).

Table 5.33. : *One-way ANOVA test results for differentiation of the SUS-TR scale based on the University' SIS Usage Frequency (* Significance Level at %95 ($P<0.05$))*

	University' SIS Usage Frequency	n	$\bar{x}$	s.d.	F	p
SUS-TR	I use it everyday.	28	57.50	8.47	1.359	0.259
	I use it few times a week.	208	58.22	10.11		
	I use it few times a month	88	60.03	7.99		

Specifically, the impact of the usage of SIS at the University over the scores of the participants has been tested via the below-described hypothesis test,

H₀: There is no statistically significant difference between scores of participants' SUS-TR scale according to SIS usage frequency at the University.

H₁: There exists a statistically significant difference between scores of participants' SUS-TR scale according to SIS usage frequency at the University.

According to the results of the one-way ANOVA test, presented in the following table, the scores of participants of the SUS-TR scale were compared with University' SIS most liked features and it was determined that the most liked features of SUS-TR scores of the University SIS was not statistically significant at 95% confidence level ($F=0.594$; $p > 0.05$) (See Table 5.34).

Table 5.34. : *One-way ANOVA Test Results for Differentiation of SUS-TR Scale according to the Most Liked Features of SIS at the University (* Significance Level at %95 (P < 0.05))*

	University'SIS Most Liked Features	n	$\bar{x}$	s.d.	F	p
SUS-TR	None	156	58.77	9.47	0.594	0.667
	Communication	25	55.90	7.53		
	Ease of use	46	58.75	8.16		
	Content	75	59.10	11.06		
	Other	22	59.20	8.11		

The impact of the favorite features of SIS at the University on the participant scores has been conducted with the following hypothesis test.

H₀: There is no statistically significant difference between participants' scores of the SUS-TR scale according to the most liked features of the University SIS.

H₁: Participants' scores of the SUS-TR scale differ significantly according to the most liked features of the University SIS.

According to the results of the one-way ANOVA test in order to test whether statistical significance of the SUS-TR scores of the participants in the survey according to the most disliked features of the university'SIS was significant, the SUS-TR scale scores were not statistically significant at 95% confidence level according to the most disliked features of the University' SIS (F = 1.630; p > 0.05) (See Table 5.35).

Table 5.35. : *One-way ANOVA Test Results for Differentiation of the SUS-TR Scale according to the Most Disliked Features of SIS at the University (* Significance Level at %95 (P < 0.05))*

	Most Disliked Features of SIS at the University	n	$\bar{x}$	s.d.	F	p
SUS-TR	None	25	61.30	8.69	1.630	0.097
	Communication	20	60.63	9.49		
	Incompatibility	15	59.33	8.99		
	Content	19	60.53	6.95		
	Functionality	52	57.31	9.05		
	Login	48	60.68	10.81		
	Accessability	62	56.13	9.12		
	Everything	16	56.25	5.16		
	Design	16	61.09	11.58		
	Acceptance	16	61.25	12.78		
	Complexity	35	56.79	8.28		

Again for the SIS at the University, how the participants' scores are affected by the most disliked features was evaluated by using the following test,

H₀: The difference of participants' SUS-TR scores according to the most disliked features of SIS is not statistically significant.

H₁: The differences of participants' SUS-TR scores according to the most disliked feature of SIS are statistically significant.

In order to test whether there exists a statistically significant difference among the scores of SUS-TR scale according to the recommendations to increase the usability of SIS at the University, the one-way ANOVA test was conducted and its results were given in Table 5.28. According to the obtained results of one-way ANOVA, SUS-TR scores were not statistically significant 95% confidence level ($F=0,867$; $p > 0,05$) (See Table 5.36).

Table 5.36. : *One-way ANOVA Test Results of Differentiation of the SUS-TR Scale according to Recommendations about Increase Usage of SIS at the University(* Significance Level at %95 (P<0.05))*

	Recommendations about Increase Usage of SIS at the University	n	$\bar{x}$	s.d.	F	p
SUS-TR	None	34	57.28	9.78	0.867	0.545
	Ease of use	27	56.48	8.61		
	Content	14	58.57	8.19		
	Login	11	56.14	6.83		
	Development	121	58.45	9.36		
	Design	11	57.50	12.40		
	Acceptance	32	59.69	9.46		
	Incompatibility	52	61.06	9.88		
	Other	22	59.20	9.86		

The impact of above-tabulated recommendations on the increasing usage of SIS was tested with the following hypothesis test,

H₀: Participants' scores of SUS-TR scale are not statistically significant with respect to the recommendations to increase the usage of SIS at the University.

H₁: Participants' scores of SUS-TR scale differ statistically with respect to the recommendations to increase the usage of SIS at the University.

According to following table, there exist a statistically significant relationship between the age groups of the participants and the SUS-TR score groups ($p: 0.046 < 0.05$). To illustrate, in total 60 participants in the 21-23 age range have OK/Fair group based on SUS Score. In addition, it was found that overall SUS-TR score of the 324 participants was 46 which was between poor and Ok/Fair. This value also belongs to an unacceptable group (See Table 5.37).

Table 5.37. : *Determination of the Difference Between Age and SUS Score Groups*

		SUS-TR Score Groups						Total
		Worst Imaginable	Awful	Poor	OK/Fair	Good	Excellent	
Age	18-20	5	25	35	26	6	0	97
	21-23	2	28	74	60	10	2	176
	24+	6	9	17	15	3	1	51
Total		13	62	126	101	19	3	324

$X^2:18,609$, $sd:10$, $p:0,046<0,05$

According to the results presented in the following table, there was a statistically significant relationship between participants' SUS Score groups and the most disliked features of SIS at the University ($p: 0.036 < 0.05$). For instance, participants in Poor subgroup (totally 26 people) based on SUS Score have the Functionality as a most disliked feature of SIS at the University (See Table 5.38).

Table 5.38. : *Determination of the Differences Between Participants' SUS-TR Score Groups and Most Disliked Features of SIS at the University*

		Participants' SUS Score Groups						Total
		Worst Imaginable	Awful	Poor	OK/Fair	Good	Excellent	
Most Disliked Features in the University' SIS	None	0	6	5	11	3	0	25
	Communication	0	5	5	7	3	0	20
	Incompatibility	1	1	4	9	0	0	15
	Content	0	2	7	10	0	0	19
	Functionality	4	7	26	13	2	0	52
	Login	1	7	19	14	6	1	48
	Accessibility	6	13	24	18	1	0	62
	Everything	0	3	12	1	0	0	16
	Design	0	3	6	5	1	1	16
	Acceptance	0	4	5	5	1	1	16
Complexity		1	11	13	8	2	0	35
Total		13	62	126	101	19	3	324

$X^2:69.385$, $sd:50$, $p:0.036<0.05$

Bangor, Kortum, and Miller (2009) isolated individual SUS scores into 7 classes. A seven-point descriptive word tied down Likert scale was added as an eleventh inquiry to about 1,000 SUS overviews. Results demonstrate that the Likert scale scores relate to a great degree well with the SUS scores ($r=0.822$). The expansion of the descriptor rating scale to the SUS may enable specialists to decipher singular SUS scores and help in disclosing the outcomes to individual components experts. In this investigation, it has been utilized an indistinguishable strategy so from to ordered the SUS-TR scores. When participants were classified according to SUS scores, 4% of participants were worst imaginable, 19.1% of those was awful, 38.9% of those was poor, 31.2% of those was Ok/Fair, 5.9% of those was good and 0.9% of those are excellent based on SUS scores.

According to the SUS-TR scores, except between the age groups and the most disliked characteristics in the SIS, there was no significant correlation other variables as a result of Chi-Square test ($p > 0.05$) (See Table 5.39).

Table 5.39. : *Participants' Categorized SUS-TR Scores*

SUS-TR Scores	Number	Percent (%)
Worst Imaginable	13	4.0
Awful	62	19.1
Poor	126	38.9
Ok	101	31.2
Good	19	5.9
Excellent	3	.9
Total	324	100.0

5.8. Relation Between EWPL-TR and SUS-TR

The relationship between the subscales of EWPL and SUS-TR measurements is examined in the following table by using Pearson Correlation Coefficient.

When the relationship between EM subscales and SUS-TR scale is examined; there was no significant relationship between EM positive and SUS-TR ($p > 0.05$, $p = 0,818$, $r = -0,013$). It has been determined that there is a moderately strong relationship between EM positive and EM negative in the positive direction ($p < 0.01$, $p = 0,000$, $r = 0,605$). There was no significant relationship between SUS-TR and EM negative. ($p > 0.05$, $p = 0.948$, $r = 0.004$) (See Table 5.40).

Table 5.40. : *Pearson Correlation Analysis for the Dependence between the subscales of EWPL and SUS-TR measurements*

	EM Positive	SUS-TR	EM Negative
EM Positive	1	-,013	,605**
SUS-TR		,818	,000
EM Negative		1	,004
			,948
			1

5.9. Discussion and Conclusion

This study was created to test the research question that is “How do different student groups’ background affect users’ emotion and perceived usability of SIS”.

Firstly, those who are registered in a major that is related with computer field have stronger emotional perceptions than those registered in a major that is unrelated with computer field. This finding is meaningful since students who are registered in a major that is related with computer field are more capable and familiar with any kind of IS. Therefore, they have stronger feelings about the SIS that was used in the study.

The study also demonstrated that students who use Internet web services rarely have higher positive emotional perceptions than those who use such services daily. Students who use Internet web services daily can be seen more experienced with any kind of IS so that

students who use Internet web services a few days in a week may be more tolerated to any kind of system error, problem, inadequacy, etc. This finding is in line with the finding that students who use the SIS rarely have lower negative emotional perceptions than those who use the SIS more frequently.

Findings also showed that students who are more experienced with the use of computers have lower positive emotional perceptions compared to those who have less experience with the use of computers. This finding is in line with the finding that students who have advanced computer usage skills have lower positive emotional perceptions than those who have intermediate and intermediate-advanced computer usage skills.

Contrary to expectations, students who feel unconfidence about their computer usage have higher negative emotional perceptions than those who feel confident about their computer usage. This finding can be explained with a relation between self-confidence and anxiety. Students can feel more anxiety when they do not feel confident with a usage of the SIS.

According to findings of the study, students who are in their fifth year in the university have lower positive emotional perceptions than those who are in their first, second and third year in the university. This finding also can be explained with fifth-year students' experience with the SIS. In addition, students that have GPA between 3.50 and 4.00 have lower positive emotional perceptions than those who have lower GPA's. It was expected that successful students will have stronger emotions.

The results also showed that students who do not like any feature related with the SIS have lower positive emotional perceptions compared to those who like features such as communication, ease of use, content and features in the other category. Contrary to expectations, students who do not like any feature related with the SIS have lower negative

emotional perceptions compared to those who like features such as ease of use, content and features in the other category. In line with the expectations, students who do not dislike any feature related with the SIS have higher positive emotional perceptions compared to those who dislike features such as communication, incompatibility, content, functionality, login, accessibility, design, acceptance and in the category of everything. Besides, students who dislike everything related with the SIS have lower positive emotional perceptions compared to those who only dislike features such as communication, incompatibility, content, functionality, login, accessibility, design, acceptance and complexity. On the other hand, contrary to expectations, students who do not dislike any feature related with the SIS have higher negative emotional perceptions compared to those who dislike features such as incompatibility, content, functionality, login, accessibility, acceptance, complexity and in the category of everything. Again contrary to expectations, students who dislike everything related with the SIS have lower negative emotional perceptions compared to those who only dislike features such as communication, incompatibility, content, functionality, login, accessibility, design, acceptance and complexity. As expected, students who do not have any recommendations to increase usage of the related SIS have higher positive emotional perceptions compared to those who have suggestions on features such as development and incompatibility. The results also showed that students who have recommendation on the design category to increase usage of the related SIS have higher positive emotional perceptions compared to those who have suggestions on features such as development and incompatibility. Besides, students who have recommendation on the development category to increase usage of the related SIS have lower negative emotional perceptions compared to those who have suggestions on features such as content, design, features in the other category and those who have no suggestions at all.

To sum up findings related with students' emotional reactions; the study showed that students' emotional reactions to the SIS differ with their demographic backgrounds. Besides, students' emotional reactions to the SIS differ based on their dis/like features or recommendations related with the SIS.

When it comes to findings about SUS-TR scores of the participants, the mean SUS-TR scores belongs to "Not Acceptable" class which is between "Poor" and "OK".

There exist a statistically significant relationship between the age groups of the participants and the categorized SUS-TR score groups. To illustrate, in total 60 participants in the 21-23 age range have OK/Fair group based on SUS Score. There was a statistically significant relationship between participants' categorized SUS-TR score groups and the most disliked features of SIS at the University. For instance, participants in poor sub-group (totally 26 people) based on SUS score have the functionality as a most disliked feature of SIS at the University. When participants were classified according to SUS scores, 4% of participants were worst imaginable, 19.1% of those was awful, 38.9% of those was poor, 31.2% of those was Ok/Fair, 5.9% of those was good and 0.9% of those are excellent based on SUS scores. According to the categorized SUS-TR scores, except between the age groups and the most disliked characteristics in the SIS, there was no significant correlation other variables as a result of Chi-Square test.

When the relationship between EM subscales and SUS-TR scale is examined; there was no significant relationship between EM positive and SUS-TR. It has been determined that there is a moderately strong relationship between EM positive and EM negative in the positive direction. There was no significant relationship between SUS and EM negative.

Therefore, analyzing users' emotional variations amongst different student groups is valuable because focusing on these variations has the potential to raise user experiences of

users with any SIS. Thus, SIS developers should be attentive of these variations and student' demographic information to create successful SIS.



6. PILOT STUDY

6.1. Introduction

Although, SIS are widely studied in the academic world, these studies showed that these systems should be still worked on to make them more productive. In order to fill this gap, selected university's SIS was evaluated in terms of user emotions, performance and perceived usability. Nielsen (1993) stated that no usability test should be performed without first having tried out the test procedure on a few pilot subjects. To our knowledge, no comprehensive usability test of selected university's SIS have been previously done. Therefore, a pilot study was conducted to identify possible problems before experimenting with large student groups.

6.2. Study Context

6.2.1. Organizational Setting

Tested system was one of the universities' SIS in Turkey. The university has over twenty thousand students. Furthermore, this SIS is only the system that is used by the university's students. Through this system, the students can get information about announcements about their departments, manage their personal data, review their academic performance as well as course histories. The SIS provides also course enrollment and communication between students and their advisors. Tested system is used by all departments' students as well as academics.

6.2.2. Objective

Firstly, participants were asked to perform a number of tasks on the tested system. After that, they were asked to answer EWPL-TR and SUS-TR to get their emotion and usability perceptions about the system, respectively. Then, performance evaluation and demographic questionnaires were carried out. Quantitative and qualitative data were collected

from only six students since the main goal of this pilot experiment was to identify possible problems before experimenting with large student groups.

6.3. Method

6.3.1. Study Design

Participants completed one set of tasks based on an artificial scenario in a laboratory setting. Quantitative and qualitative analyses, including one list of artificial scenario, SUS-TR, EWPL-TR, performance evaluation test, demographic information and overall assessment questionnaires were applied to six students. In this study, the SIS user interface was displayed on 15.6 inch WLED monitor with 1920 by 1080 pixel resolution using Windows 10 via Google Chrome browser. All students were participated to the study in computer laboratory of the Center for Information Systems at the university. There were one facilitator (one of the authors of this thesis) and a participant at each case of the experiment. Toward the start of the assessment session, the member marked an educated consent form, recognizing that: participation is intentional, and that volunteers have the privilege to quit participation whenever. The members were then requested to look the printed assignment description and urged to verbally process as they finished each assignment. The facilitator observed the participant and also guided them throughout the session.

6.3.2. Participants

There is a debate how much participant is needed to conduct a usability test. It is widely assumed that five participants enough for usability studies. Faulkner (2003) informed that some of the randomly selected of five participants found 99% of the usability problems. In this study, pilot test was conducted to six participants. Aged of participants ranged from 21 to 26. There were three females and three males. Besides, there were five undergraduate students and one graduate student. Their majors were International Finance, International

Trade and Business and Economy. Their graduate point averages (GPA) are as the following; two of the participants' GPA were between 3.00 and 3.49, one of the participant's GPA was between 2.50 and 2.99, two of the participants' GPA were between 2.00 and 2.49 and one of the participant's GPA was under the 2.00. Four of the participants expressed their computer skills as intermediate level and two of the participants described their computer skills between intermediate and advanced levels. One of the participants mentioned that s/he is confident about using the computer and five of the participants stated that their confidence in the use of computers could change depending on the given task. All of the participants indicated that they are using the Internet everyday and their daily computer usage ranged from one hour to eight hours. Lastly, three of the participants stated that they are using the tested SIS several times a week and the other three of the participants stated that they are using the tested SIS several times a month.

6.3.3. Test Scenarios

Although both academics and students of the university use the SIS application, they have different academic responsibilities and therefore use different templates provided by the application. Since only this study was performed to get students perspective about the system, only academic tasks commonly carried out by them were selected. One list of artificial test scenarios was given each participant. These tasks will be shown and detailed in the following chapter (See Chapter 7).

6.3.4. Usability Analysis

Tested system's usability was evaluated through questionnaires, performance measures and participants' overall assessment responds.

6.3.5. Outcome Measures

6.3.5.1. EWPL-TR

Firstly, professionally translated Emotion Word Prompt List's reability and validity studies have been carried out and Turkish version of the scale was finalized which is the EWPL-TR (See Chapter 3). Participants' urgent emotional responses to the system were measured using a 15 item scale of emotion words commonly used to describe reactions to web sites which is the EWPL-TR. EWPL-TR consists of 9 positive, 6 negative word.

EWPL-TR was accomplished by each participant at the end of testing session.

6.3.5.2. Performance measures

Two performance metrics used in the study:

1. Task success rate : It evaluates the percentage of a given assignment that participants successfully complete without any important errors. These errors result in an incorrect or incomplete outcome.
2. Percentage of correct answers: A questionnaire that consists of 10 questions and a 5-point choices was conducted to measure the user performance.

6.3.5.3. SUS-TR

In addition to emotional and performance measures, the system usability scale (SUS) was conducted to get participants' perceived usability on the system.

SUS has been developed in the language of English so its translation, reliability and validity studies have been carried out before the pilot study (See Chapter 4).

The SUS-TR was completed by each participant at the end of testing session.

6.4. Results

All participants were completed the evaluation sessions without any technical difficulties. Although number of participants of the pilot study is too small to generalize findings, results give some evidence that participants' emotional reactions to SIS, and likes, dislikes and suggestions related with the aspects of the SIS differ based on their demographic information. To illustrate, female participants showed more negative emotions to SIS compared to male participants. Therefore, analyzing users' emotional variations amongst different demographical groups is valuable because focusing on these variations has the potential to raise efficiency of user experiences with SIS.

It can be also expected that users' ratings of the usability of the SIS may differ depending on their demographic information. To illustrate, students with high GPA gave lower usability scores to SIS compared to students with low GPA.

Discoveries also gave prompts about other important discoveries. For instance, positive feelings may bring high usability appraisals of a SIS and negative feelings may bring low usability evaluations for the SIS. Thus, focusing on users' emotional responses has a potential to raise SIS' usability ratings. Participants' overall SUS score was 58.5 which is marginal.

Participants' task success rate ranged from six to eight. Three of the participants task success were seven (88%), two of the participants' task success were six (75%) and one of the participants completed all of the eight tasks (100%). Participants' number of the correct answers ranged from five to 10. One of the participants replied five correct answers (50%), one of the participants replied six correct answers (60%), two the participants replied seven correct answers (70%), one of the participants replied nine correct answers (90%) and one of the participants replied 10 correct answers (100%). Although the number of participants are

too small to perform analysis to get correlations between variables, it can be predicted that user performance have correlations with both user emotions and users' perceived usability ratings.

6.5. Discussion

It has previously been shown that the adoption and advanced system implementation of SIS has the potential to enhance the efficiency and quality of academic workflow. This pilot study sets ground for future usability studies for SIS. To generalize the outcomes from the pilot study, further study needs to be done with a wider range of users. With the wider data, more analysis can be done on users' emotional reactions, performance, perceived usability and demographic information. Furthermore, with a large group it may be possible to use students' demographic information to predict their preferences for some specific features on SIS.

6.5.1. Limitations of the study

There were some essential methodological impediments in this examination. This investigation incorporates an usability test with a predetermined number of assignments in a solitary foundation as they finished a set number of scholastic tasks. Additionally, none of the usability software is used in this study. Thusly, it has not been captured important usability evaluation datas such as time spent on each task, mouse clicks, mouse movements, web page changes, onscreen text, error types and counts. In addition, eye tracking data may give more accurate clues about usability of the tested system.

6.6. Conclusion

The main goal of this pilot experiment was before experimenting with large student groups to identify possible problems. Furthermore, there is no scientific evidence on usability of any Turkish Universities' SIS in the literature. So that, this study showed that this study's method can be used to evaluate students' emotions, performance and perceived usability on a

specific SIS. This pilot study will serve as a baseline study for a future comparative usability evaluation of SIS in other institutions. This method can be applied to the evaluation of other academic applications as well. In order to obtain more significant results and to generalize findings, extended studies will include larger groups of students (See Chapter 7).



7. EVALUATION OF STUDENT INFORMATION SYSTEM (SIS) IN TERMS OF USER EMOTION, PERFORMANCE AND PERCEIVED USABILITY

7.1. Introduction

As discussed in the literature review (see subsection 2.4), there have been a few efforts to investigate the effects of user experience and perceived usability on SIS (Battleson, Booth, & Weintrop, 2001; Carcary, Long & Remenyi, 2007; Gemmell & Pagano (2003)). In the light of findings from these researches, it is obvious that there is value in investigating the evaluation of SIS in terms of user emotions, performance and perceived usability. Therefore, as explained in the previous chapters since our users' main language was Turkish, SUS and EWPL questionnaires have been translated and reliability, validity and pilot studies have been carried out in order to perform this experimental usability study (See Chapter 4, Chapter 5 and Chapter 6).

In summary, this study investigated whether students' demographic information affect their emotion, performance and perception of the usability of the studied SIS.

7.2. Method

7.2.1. Design

Different methods have been administered in order to examine the usability of selected University's SIS. A test system status was used in order not to affect the real records and the usability tests were conducted by using this status. The selected university is one of the foundation universities in Turkey- Istanbul city. The selected university has over 20.000 students however the name of the university will not be mentioned because of private reasons.

Thereafter, the usability tests were performed with two groups of real system users. A screen recording device was used during the usability studies to record and monitor the screen

and calculate the click numbers of the participants. Task analysis was performed so as to decide the basic tasks and to illustrate how these tasks are accomplished step by step.

The participants were chosen among the University's first cycle students. The first group is prone to technology departments which are Management Information Systems and Information Systems & Technologies students. The second group is from the rest of Faculty of Commerce departments. A total of 16 participants from each group were included in the study. In each group, There were two participants, one female and one male in the freshman; two participants, one female and one male in the sophomore; two participants one female and one male in the junior and two participants, one female and one male in the senior (See Figure 7.1 and 7.2).

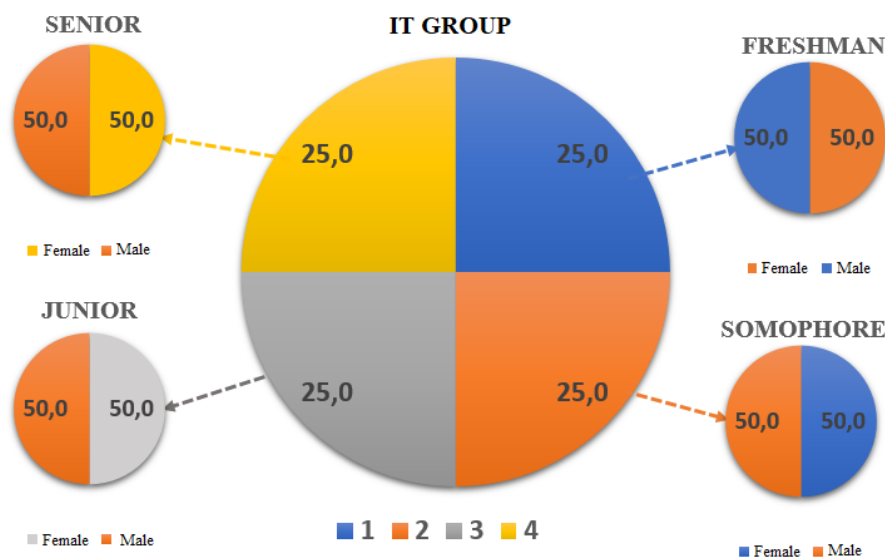


Figure 7.1. :A sample of the IT Group Population

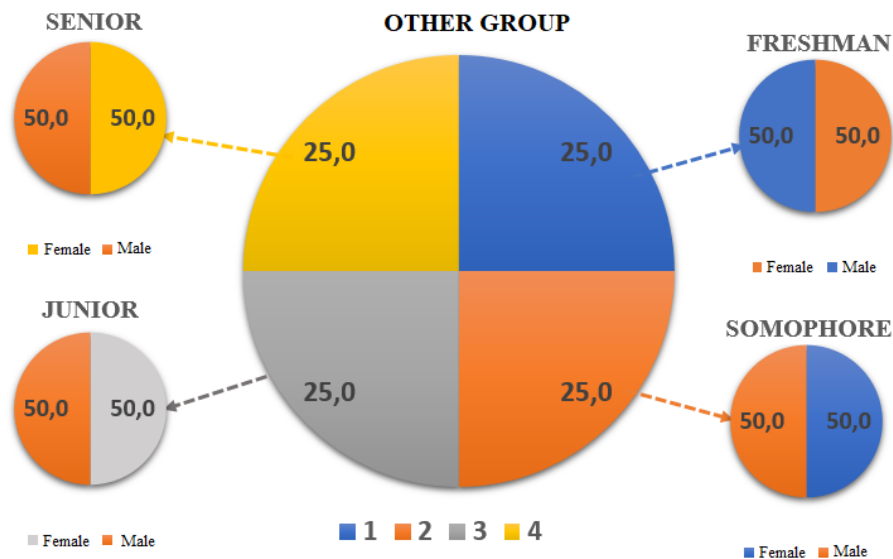


Figure 7.2.: A sample of the Other Group Population

Every member attempted a progression of tasks on the chose SIS. These assignments were deliberately developed to control members to components of the SIS that would understudies experience issues in completing. In the wake of finishing the assignments on the system, they finished a progression of short surveys (a) their prompt emotional response to the system (as estimated by the EWPL-TR); (b) their impression of the usability of the system (as estimated by the SUS-TR); (c) their like/dislike on the system (as estimated by the Overall Reactions Questionnaire, ORQ). Also, the click counts and time spent on each task within the selected SIS were calculated. Test sessions conducted with the users were recorded to measure how much time each task took and to observe whether they focused on the relevant part of the screens or objects.

Thus there were three groups of dependent variables:

- (a) User experience (assessed by using the EWPL-TR and the ORQ)
- (b) Performance (assessed by the time spent on each task within the SIS, users' click counts on tasks and number of successful tasks by carried out participants)

(c) Perceived usability (assessed by the SUS-TR)

7.2.2. Participants

In the following table, when the demographic distributions of the survey participants were examined; it has been observed that 50% were male, 50% were female and 29.9% of the age groups were 18-20, 54.3% were 21-23, 15.7% were over 24 years old (See Table 7.1).

Table 7.1. : *Demographic Information of the Working Group*

		f	%
Gender	Female	16	50
	Male	16	50
Total		32	100
Age	18-20	11	29.9
	21-23	16	54.3
	24+	5	15.7
Total		32	100

Besides, 12.5% of participants were in the first year , 15.6% of them were in their second year, 28.1% of them were in their third year, 18.7% of them were in their fourth year, 18.7% of them were in their fifth year and 6.2% of them were in their sixth year and over (the time they spent at the university was determined). Based on the separate fields, 50% of the participants were in the departments that are related to computer field and 50% of the participants were in the departments that are unrelated to the computer field.

When the academic grade point average (GPA) was examined for those participants, the data showed that 18.7% of the participants' GPA was below 2.00, 18.7% of the participants' GPA between 2.00 and 2.49, 25% of the participants' GPA between 2.50 and 2.99, 18.7% of the participants' GPA between 3.00 and 3.49, and 6.25% of the participants' GPA between 3.50 and 4.00, and 12.5% of the participants did not provide their GPA (See Table 7.2).

Table 7.2. : *Participants' Educational Demographics*

Education Level	n	%
Undergraduate	32	100.0
Total	32	100.0
Duration at the University?	n	%
1	4	12.5
2	5	15.6
3	9	28.1
4	6	18.7
5	6	18.7
6+	2	6.2
Total	32	100.0
Registered department	n	%
Computer field (Information Systems and Technologies,MIS, E-commerce)	16	50.0
Other	16	50.0
Total	32	100.0
Current GPA	n	%
Below 2.00	6	18.7
Between 2.00 and 2.49	6	18.7
Between 2.50 and 2.99	8	25
Between 3.00 and 3.49	6	18.7
Between 3.50 and 4.00	2	6.25
I do not know.	4	12.5
Total	32	100.0

Besides, 21.8% of the participants have been using the computer for 5-7 years and 78.1% of the participants have been using the computer for 10 years and over. Furthermore, about the level of computer usage, it was determined that 46.8% of participants were at the intermediate level, 37.5% were at intermediate to advanced level and 15.6% had advanced computer skills. Additionally, while participants indicated that 9.3% would generally need help in terms of computer usage, 50.0% said that it could change depending on the given task and 40.6% said they were confident in terms of computer usage.

When the participants were analyzed according to their daily average computer usage time, 75.0% of the participants stated that they use their computers 1-3 hours per day, 21.8% of them stated that they use their computers for 4 hours and above. Based on the frequency of use of Internet-Web services by participants, it was examined that participants use 96.8% every day and 7.4% use a few days a week. As a more specific character for the participants, the frequency of using the SIS was also examined and it was determined that 18.7% of the participants use the student information system every day, 50.0% use a few days a week and 31.2% use SIS for a few days a month (See Table 7.3).

Table 7.3. : *Computer Usage Characteristics of the Participants in the Experimental Study*

Computer usage duration	n	%
I use it for 5-7 years.	7	21.8
I use it for 7-10 years.	0	0.0
I use it for 10 years and over.	25	78.1
Total	32	100.0
Computer skills	n	%
Intermediate level	15	46.8
Intermediate to Advanced	12	37.5
Advanced	5	15.6
Total	32	100.0
Self-confidence about computer usage	n	%
I usually need help.	3	9.3
It depends on the given task.	16	50.0
I trust.	13	40.6
Total	32	100.0
Daily computer usage	n	%
1-3 hours	24	75.0
4 hours and over	7	21.8
Total	32	100.0
Internet and web-services usage frequency	n	%
I use it every day.	31	96.8
I use it a few days a week.	1	3.1

Total	32	100.0
SIS usage frequency	n	%
I use it every day.	6	18.7
I use it a few days a week.	16	50.0
I use it a few days a month.	10	31.2
Total	32	100.0

To analyze the details of SIS, various questions have been asked to participants in detail. To illustrate, when participants asked about the most favourite features in SIS at the University,

34.3% those agreed with ease of use, 15.6% of the participants said none, ie. no feature is selected as a favourite one. Besides, 15.6% of those agreed with communication, 15.6% of those agreed with functionality, 9.3% of the participants said that accessibility, 6.2.% those agreed with content and 3.1% of the participants said that login for a favorite feature of SIS.

After all, the most mentioned category is ease of use and the least mentioned category is login.

In a similar manner, when the most disliked characteristic of SIS at the University was examined; 21.8% of participants agreed with functionality, 15.6% of those agreed with login, 15.6% of those agreed with acceptance and 15.6% of those defined complexity as a disliked feature, 12.5% of those agreed with accessibility, 6.2% of the participants declared none, which means no disliked feature at all. Furthermore, 3.1% of those participants accepted communication as a disliked feature and 3.1% of those agreed with incompatibility, 3.1% of those agreed with content, 3.1% of those agreed with design, as it is given in the following table

After all, the most mentioned category is functionality and the least mentioned categories are communication, incompatibility, content and design.

Finally, the necessary improvements to be done in order to increase the usability of the SIS were examined, it was determined that 31.2% of the participants mentioned incompatibility, 18.7% of those mentioned ease of use, 9.3% of those mentioned development, 9.3% of those mentioned as communication, 6.2% of the participants selected as none means there is no need to change, 6.2% of the participants mentioned content, 3.1% of the participants mentioned design, 3.1% of the participants mentioned acceptance, and finally 3.1% of the participants mentioned other.

After all, the most mentioned category is incompatibility and the least mentioned categories are design, acceptance and other (See Table 7.4).

Table 7.4. : Participants' Categorized Overall Reactions Questionnaire Responses

The most liked features in the University'SIS	n	%
None	5	15.6
Communication	5	15.6
Ease of use	11	34.3
Accessibility	3	9.3
Functionality	5	15.6
Content	2	6.2
Login	1	3.1
Total	32	100.0
The most disliked features in the University'SIS	n	%
None	2	6.2
Communication	1	3.1
Incompatibility	1	3.1
Content	1	3.1
Functionality	7	21.8
Login	5	15.6
Accessibility	4	12.5
Everything	0	0
Design	1	3.1
Acceptance	5	15.6

Complexity	5	15.6
Total	32	100.0
Recommendation to improve the usability of the University'SIS	n	%
None	2	6.2
Ease of use	6	18.7
Content	2	6.2
Login	0	0
Development	3	9.3
Design	1	3.1
Acceptance	1	3.1
Incompatibility	10	31.2
Accessibility	3	9.3
Communication	3	9.3
Other	1	3.1
Total	32	100.0

7.2.3. Equipment

In this study, the SIS user interface was displayed on 15.6-inch WLED monitor with 1920 by 1080 pixel resolution using Windows 10 via Google Chrome browser. All students participated in the study in computer laboratory of the Center of Information Systems at the university. There was one facilitator (one of the authors of this thesis) and a participant in each case of the experiment. Toward the start of the assessment session, the member marked an educated consent form, recognizing that: support is willful, and that volunteers have the privilege to quit participation whenever. The members were then requested to look the printed assignment list and urged to verbally process as they finished each assignment. The facilitator observed the participant and also guided them throughout the session. The Camtasia Studio 9 program was developed before the test to calculate mouse click numbers and the duration of each test.

7.2.4. SIS and Tasks

Participants were requested to complete a set of tasks on the selected University'SIS. Although both academics and students of the university use the SIS application, they have different academic responsibilities and therefore use different templates provided by the

application. Since only this study was performed to get students perspective about the system, only academic tasks commonly performed by them were selected. In addition expert opinion who work Information Technology department in the University has been received. One list of artificial test scenarios was given to each participant. These tasks are shown in the following table (See Table 7.5). The Turkish version of the artificial tasks was given in the appendix (See Appendix Table C.1.).

Table 7.5. : *Artificial Tasks and Test Scenarios*

Task Number	Tasks
1	Log in with your username and password.
2	You want to access the transcript. Please provide the page you need to use to perform this action.
3	You want to send a message to your academic advisor through the system. Please provide the page you need to use to perform this action.
4	You want to access help videos about the system. Please find out how to follow this path.
5	You want to change your login password. Please find out how to follow this path.
6	You want to choose a course from your department. Please provide the page you need to use to perform this action.
7	You want to choose a course that is not associated with your department. Please provide the page you need to use to perform this action.
8	You want to see the deadline for the financial approval. Please provide the page you need to use to perform this action.

7.2.5. Questionnaires

Participants completed an online version of EWPL-TR, SUS-TR, ORQ and demographic questionnaires about their experience of the website, delivered via SurveyMonkey, an online survey package.

7.2.5.1. Voluntary participation form

It is intended to give information about the subjects to participate in the research, the purpose of the test and its general structure, to define the rights of the subjects and to obtain approval for participation in the test. It has been noted that the information contained in the consent form is brief and informative, and the personal information of the subjects is not included on the form (See Appendix C.2.)

7.2.5.2. EWPL-TR

Participants' urgent emotional responses to the system were measured using a translated 15 item scale of emotion words commonly used to describe reactions to websites, the Emotional Words Prompt List (EWPL-TR) (See Table B.2 in Appendix).

7.2.5.3. SUS-TR

The aim of asking about usability was to see whether people with demographic background rated the usability of the SIS differently. Participants' perceived usability was measured by the translated SUS-TR (See Table B.4 in Appendix).

7.2.5.4. Overall Reactions Questionnaire (ORQ)

Participants were asked open-ended questions about the system to see whether they mentioned things that match their emotion and perceived usability. Participants were asked what they most liked and most disliked and suggested the increased usability of SIS about the system (See Appendix C.1.).

7.2.5.5. Demographic Questionnaire

As a last step of data collection, participants completed a demographic questionnaire that gathered information on age, sex, GPA, registered department, computer usage confidences, internet - web service and SIS usage frequencies (See Appendix C.4.)

7.2.6. Procedure

The procedure for each participant was:

Start the logging application on one of the computers in the computer-lab

Accomplish the voluntary participation form.

Accomplish the selected university's SIS tasks.

Accomplish the Emotional Word Prompt List (EWPL-TR) for the SIS.

Accomplish the System Usability Scale (SUS-TR) for the SIS.

Accomplish the Overall Reactions Questionnaire (ORQ) for the SIS.

Accomplish the demographic questionnaire.

7.3. Data Analysis

32 people participated in the study.

In the analysis of data SPSS Statistics Solution has been utilized. Nonparametric tests were used in the analysis of data with no normal distribution. Mann-Whitney U in the mean comparison of 2 independent groups; Kruskal-Wallis tests were used to compare the mean of two independent groups.

The next subsections will demonstrate data preparation on each variable as required.

7.3.1. Student Background

There were sufficient participants with valid data from different groups to create two groups of students. There was 16 computer field (Information Systems and Technologies, Management Information Systems) and 16 other registered department students.

7.3.2. Emotion Word Prompt List (EWPL-TR) Data

Ratings on the 15 emotional words (9 positives and 6 negatives) in the EWPL-TR were coded from 1 to 5 (See Table A.2. in Appendix). Earlier, detailed information was given about EWPL-TR (See Chapter 3).

7.3.3. System Usability Scale (SUS-TR) Data

Ratings on the 10 words (5 positives and 5 negatives) in the SUS-TR were coded from 1 to 5 (See Table A.4. in Appendix). Earlier, detailed information was given about SUS-TR (See Chapter 4).

7.3.4. Overall Reactions Questionnaire (ORQ) Answers

Participants' answers to overall reactions to the SIS consisted of free text. For that reason, content analysis was performed to analyze this free text and a set of categories of reactions was extracted. Beforehand, detailed information was given about this procedure and content analysis categories (See Chapter 5).

7.3.5. Effects of Different Student Groups on Users' Emotion, Performance and Perceived Usability on SIS

Effects of different student groups on users' emotion, performance and perceived usability on SIS will be presented in this sub-section.

When the mean and standard deviations of the EWPL-TR and SUS-TR scales of IT and OTHER group were examined in the following table, it was determined that the mean value of EM positive of the IT group was 25.56 ± 8.87 and the mean value of EM negative was 15.50 ± 7.07 . For the OTHER group, EM positive and negative mean scores were found

to be 24.31 ± 8.66 and 16.56 ± 7.62 , respectively. Besides, the mean value of SUS-TR of the IT group was 61.72 ± 25.44 and the mean value of SUS-TR of the OTHER group was 61.72 ± 28.82 (See Table 7.6).

Table 7.6. : Descriptive Statistics of the Participants by EWPL-TR and SUS subscales Based on Questionnaires

GROUPS		EM Positive	EM Negative	SUS-TR	EWPL- TR	SUS-TR
IT	$\bar{X}$	25.56	15.50	61.72	40.97	61.72
	s.d	8.87	7.07	25.44		
OTHER	$\bar{X}$	24.31	16.56	61.72		
	s.d	8.66	7.62	28.82		

Mann-Whitney U was conducted in order to test whether the perceptions of participants of EWPL-TR subscales differ in terms of groups or not. Based on the obtained results in the following table, there exists no statistically significant difference between the groups under each subscale based on p values ($P=0.763$, $P > 0.05$ for EM Positive and $P=0.557$, $P > 0.05$ for EM Negative) (See Table 7.7).

Table 7.7. : Mann Whitney U Test Results of EWPL-TR Subscales according to Different Student Groups

	GROUPS	n	Mean Rank	U	P
EM Positive	IT	16	17.0	25.5	0.763
	OTHER	16	16.0		
EM Negative	IT	16	15.5	30.5	0.557
	OTHER	16	17.5		

H₀: The differences in perceptions of participants of EWPL-TR subscales are not statistically significant according to groups.

H₁: The differences in perceptions of participants of EWPL-TR subscales are statistically significant according to groups.

As a result of the obtained test result, presented in the following table., there exists a significant difference of the perceptions of the participants in OTHER group with EM Negative having different mean ranks of a most liked feature of SIS ($X^2=10.863$, $p=0.028$; $p < 0.05$). However, there is no difference for the other groups with EWPL-TR subscales. According to the Mann Whitney U test results, EM negative levels were found to be significantly different between the participants who have most favoured features are ease of use and communication. For instance, the most liked feature was found to be ease of use for the participants with Mean Rank = 7.33 compared to those who have the favoured feature is communication with Mean Rank = 6.17 (See Table 7.8).

Table 7.8. : *Kruskal Wallis H Test results for Differentiation of EWPL-TR Subscales based on the Most Liked Feature Group of SIS (* Significance Level at %95 (P<0.05))*

GROUPS	The most liked feature on the SIS	n	Mean Rank	X ²	df	p	
IT	EM Positive	Accessibility	1	9.50	6.448	6	0.375
		Ease of use	5	11.60			
		Functionality	3	7.70			
		Content	2	9.50			
		None	2	8.50			
		Communication	2	4.30			
		Login	1	1.00			
	EM Negative	Accessibility	1	2.50	7.851	6	0.249
		Ease of use	5	6.70			
		Functionality	3	10.50			
		Content	2	6.50			
		None	2	6.80			
		Communication	2	13.00			
		Login	1	16.00			
OTHER	EM Positive	Accessibility	2	13.25	9.389	4	0.052
		Ease of use	6	9.67			
		Functionality	1	11.00			
		None	4	2.50			
		Communication	3	10.17			
	EM	Accessibility	2	2.25	10.863	4	0.028*

Negative	Ease of use	6	7.33
	Functionality	1	14.50
	None	4	13.63
	Communication	3	6.17

H₀: Perceptions of participants of EWPL-TR subscales are not statistically significant with respect to groups that have the different most liked feature of SIS.

H₁: Perceptions of participants of EWPL-TR subscales are statistically significant with respect to groups that have the different most liked feature of SIS.

According to the EWPL positive and negative perceptions of the students who were in IT and OTHER departments in the study, there were no differences in according to gender, age, duration at the university, registered department, GPA, computer skills, confidence in computer use, daily average computer usage time, frequency of use of internet web services, the frequency of usage of SIS, the most disliked features of SIS and the recommendation categories presented to increase SIS use at 95% confidence interval ($p > 0.05$).

According to the result of the Kruskal Wallis H test in order to determine whether the participants' SUS scores based on the most liked features of the SIS were statistically significant, the difference of the SUS-TR scores of the IT group compared to the OTHER group was not statistically significant ($p > 0.05$) (See Table 7.9).

The SUS scores of the OTHER group were found to be statistically significant at 95% confidence level according to the most liked feature in the SIS ($X^2 = 9.357$, $p = 0.025$, $p < 0.05$). According to the Mann Whitney U test to determine the groups from which the difference originated, it was determined that the SUS scores differed significantly between who prefer accessibility and communication categories. The most liked feature was found to

be higher in those with accessibility (**Mean Rank = 14.00**) than those with communication (**Mean Rank = 10.17**).

Table 7.9. : *Kruskal Wallis H Test Results for SUS-TR Scales according to the Most Liked Feature Group in the SIS (* Significance Level at %95 (P<0.05)).*

GROUPS	The most liked feature on the SIS	n	Mean Rank	X ²	df	p
IT	Accessibility	1	14.00	10.395	6	0.109
	Ease of use	5	11.50			
	Functionality	3	8.67			
	Content	2	11.25			
	None	2	4.25			
	Communication	2	2.50			
	Login	1	2.50			
OTHER	Accessibility	2	14.00	10.753	4	0.029*
	Ease of use	6	10.42			
	Functionality	1	4.00			
	None	4	2.75			
	Communication	3	10.17			

H₀: Participants' SUS perceptions are not statistically significant compared to the most liked attribute in the SIS.

H₁: Participants' SUS perceptions are statistically significant compared to the most liked attribute in the SIS.

It was determined that the difference of the SUS-TR scores of the IT group according to the class groups was statistically significant at the 95% confidence level ($\chi^2=7.827$, $p=0.050$; $p < 0.05$). According to the Mann Whitney U test to determine the groups from which the difference originated, it was determined that there was a significant difference in scores between first and fourth-grade students. SUS scores of those with fourth grade (**Mean Rank = 5.88**) were lower than those with first grade (**Mean Rank = 12.50**) (See Table 7.10).

The Kruskal Wallis H test was used to determine whether the differences between the SUS scores of the participants according to class groups were statistically significant, the

difference between the SUS scores of the OTHER group was not statistically significant ($p > 0.05$) (See Table 7.10).

Table 7.10. : *Kruskal Wallis H Test Results for SUS Subscales According to Class Groups*

(* Significance Level at %95 ($P < 0.05$))

GROUPS		Class	n	Mean Rank	X ²	df	p
IT	SUS-TR	1	4	12.50	7.827	3	0.050*
		2	4	4.63			
		3	4	11.00			
		4	4	5.88			
OTHER	SUS-TR	1	4	9.38	0.536	3	0.911
		2	4	8.13			
		3	4	7.25			
		4	4	9.25			

H₀: There was no statistically significant difference between participants' scores of SUS-TR scale according to different classes.

H₁: The differences in participants' scores of SUS-TR scale are statistically significant according to different classes.

The SUS scores of the students in the IT and the Other group showed that the SUS scores were related to GPA, computer usage time, computer skills, confidence to use computers, daily average computer usage times and frequency of use of internet-web services, gender, most liked features of SIS, most disliked features of SIS and the recommendation categories to increase the usage of SIS was not statistically significant at 95% confidence interval ($p > 0.05$).

Based on the classification presented in Table 7.11, IT group participants, who were successful in T1, T3, T5, T6 with 100%; in T2 and T8 with 94%; in T7 with 69%; in T4 with 44%; and in average 88% of overall success. Meanwhile, when the OTHER group subjects were considered, they are successful in T1, T5 with 100%; in T6 with 94%; in T3 with 88%;

in T2 and T8 with 81%; in T7 with 69%, in T4 with 50%; and in average 83% of overall success.

Table 7.11. : *Mann Whitney U Test Results of Task Success in the System and Differences between IT and Other Group (* Significance Level at %95 ($P < 0.05$)).*

TASKS	IT GROUP		OTHER GROUP		U	P
	n ^a	%	n ^a	%		
T1	16	100	16	100	88.5	0.136
T2	15	94	13	81	75.5	0.047*
T3	16	100	14	88	109.5	0.484
T4	7	44	8	50	106.0	0.406
T5	16	100	16	100	111.5	0.532
T6	16	100	15	94	124.5	0.895
T7	11	69	11	69	113.0	0.571
T8	15	94	13	81	105.5	0.396
Mean	14	88	13.25	83	114.0	0.597

a The number of subjects who successfully completed the relevant task from a total of 16 participants in each experimental group

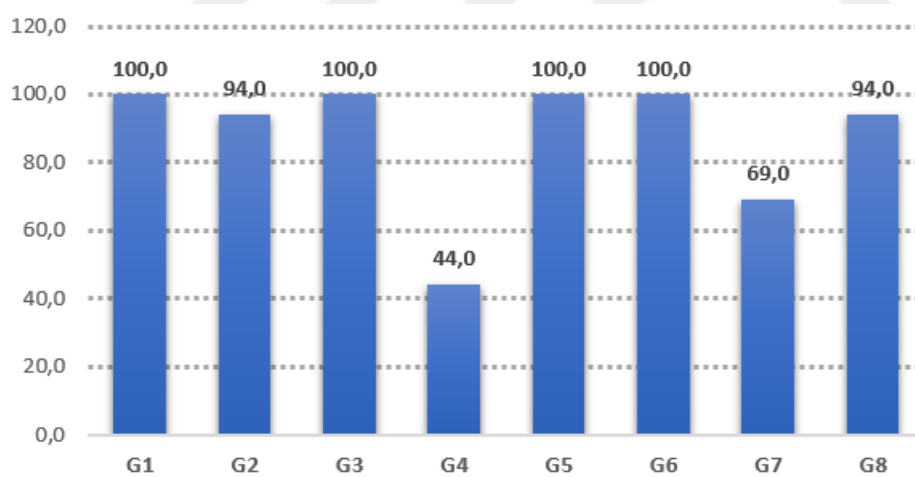


Figure 7.3. : Task Success Results for Information Technology Group Students

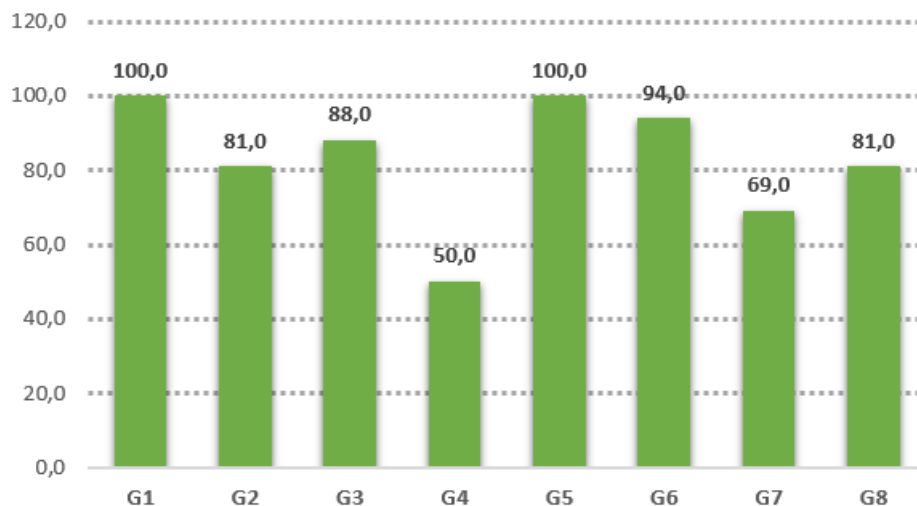


Figure 7.4. : Task Success Results for Other Group Students

Additionally, their mean rank for each task under both groups is summarized in the following table. When the differences in the time spent for the tasks according to the system used in the research were examined according to students in the IT and OTHER groups, it was determined that only the time spent for the second task (T2) was significant when it is compared to the others ($U = 75.5$; $p = 0.047$; $p < 0.05$). Furthermore, it was observed from Table 7.9. that the IT group has Mean Rank = 19.78 about the time spent for T2 and it was higher than the net time spent by OTHER group that has the Mean Rank = 13.22. Equivalently, the OTHER group has been completed T2 in a shorter time than the IT group (See Table 7.12).

Table 7.12. : Mann Whitney U Test Results of Net Times Spent on Tasks according to System Differences between IT and Other Group (* Significance Level at %95 ($P < 0.05$)).

TASKS	IT GROUP	OTHER GROUP	U	P
	Mean Rank	Mean Rank		
T1	14.03	18.97	88.5	0.136
T2	19.78	13.22	75.5	0.047*
T3	17.66	15.34	109.5	0.484
T4	17.88	15.13	106.0	0.406
T5	15.47	17.53	111.5	0.532
T6	16.72	16.28	124.5	0.895
T7	17.44	15.56	113.0	0.571
T8	17.91	15.09	105.5	0.396
General	17.38	15.63	114.0	0.597

When the number of clicks for tasks according to the site used in the survey was examined for both IT and OTHER groups, it was determined that only task 7 (T7) differs in terms of the time spent by IT group and the OTHER group (**U = 62,0; p = 0.013; p < 0.05 in Table 7.20.**) . It was determined that the IT group has a Mean Rank = 20.63 which is higher than the OTHER group, which has Mean Rank = 12.38 in terms of a number of clicks for the T7. It has been determined that the OTHER group completes the task T7 with less number of clicks compared to the IT group (See Table 7.13).

Table 7.13. : *Mann Whitney U Test Results of Number of Clicks for the Tasks according to System Differences between IT and Other Group (* Significance Level at %95 (P<0.05)).*

TASKS	IT GROUP	OTHER GROUP	U	P
	Mean Rank	Mean Rank		
T1	14.09	18.91	89.5	0.145
T2	16.28	16.72	124.5	0.892
T3	18.78	14.22	91.5	0.160
T4	18.44	13.40	81.0	0.122
T5	17.50	15.50	112.0	0.524
T6	15.53	17.47	112.5	0.540
T7	20.63	12.38	62.0	0.013*
T8	17.53	15.47	111.5	0.531
General	18.19	14.81	101.0	0.309

In the following table, the relationship between the amount of time spent in the IT group and the number of clicks was statistically significant at a 95% confidence level ($P = 0.000$; $p < 0.05$; $r = 0.832$). Similarly, the relationship between the time spent in the OTHER group and the number of clicks was statistically significant ($P = 0.000$; $p < 0.05$; $r = 0.775$). Overall, this relationship is stronger in IT group compared to the OTHER group because of the calculated correlation coefficients (See Table 7.14).

Table 7.14. : *Relationship Between the Time Spent and Number of Clicks by IT and Other Group (* Significance Level at %95 (P < 0.05)).*

GROUPS	Variables	Time Spent	Clicks
IT	Time Spent	r	1.000
		p	.000
	Clicks	r	.832**
		p	.000
OTHER	Time Spent	r	1.000
		p	.000
	Clicks	r	.775**
		p	.000

According to Pearson's correlation analysis results for determining the meaningful relationship between the EM positive and EM negative perceptions of the participants and the mean net time, mean number of clicks and mean a number of successes in the task, no relationship was detected ($p > 0.05$) (See Table 7.15).

Table 7.15. : *Pearson Correlation Analysis Results between the different dimension of the EWPL-TR perceptions of the participants and Mean Net Time, Mean number of clicks and mean number of successful tasks. (* Significance Level at %95 ($P < 0.05$)).*

		Mean Net Time	Mean Number of Clicks	Mean number of successful tasks	EM Positive	EM Negative
Mean Net Time	r	1	.755**	-0.214	0.057	-0.083
	p		0	0.241	0.758	0.652
Mean number of clicks	r		1	-0.116	0.061	-0.205
	p			0.527	0.739	0.26
Mean number of successful tasks	r			1	0.159	-0.114
	p				0.386	0.533
EM Positive	r				1	-.756**
	p					0
EM Negative	r					1
	p					

H₀: There is no significant relationship between the different dimensions of the EWPL-TR of the participants' perceptions and mean net time, the mean number of clicks and mean number of successful tasks.

H₁: There is a significant relationship between the different dimensions of the EWPL-TR of the participants' perceptions and mean net time, the mean number of clicks and mean a number of successful tasks.

According to the results of Pearson correlation analysis to determine the meaningful relationship between the total SUS-TR scores and the average net time, the number of clicks and the number of successful tasks, there was no significant relationship between the total

SUS-TR scores of the participants and the average time, the number of clicks and the number of successful tasks ($p > 0.05$) (See Table 7.16).

Table 7.16. : *Pearson Correlation Analysis Results between SUS-TR scores of the participants and Mean Net Time, Mean number of clicks and mean number of successful tasks. (* Significance Level at %95 ($P < 0.05$)).*

	Mean Net Time	Mean number of Clicks	Mean number of successful tasks	SUS-TR Scores
Mean Net Time	r	1	,755**	-0,054
	p	0,000	0,241	0,771
Mean number of Clicks	r	1	-0,116	0,121
	p		0,527	0,51
Mean number of successful tasks	r		1	0,181
	p			0,322
SUS-TR Scores	r			1
	p			

H₀: There is no significant relationship between SUS-TR total scores of the participants and mean net time, the mean number of clicks and mean number of successful tasks.

H₁: There is a significant relationship between SUS-TR total scores of the participants and mean net time, the mean number of clicks and mean number of successful tasks.

According to the results of the independent sample t-test to determine whether the participants had significant differences in terms of mean net time, mean number of clicks and mean number of successful tasks, there was no statistically significant difference between the mean net time and the mean number of successful tasks according to gender groups ($p > 0.05$).

There was the statistically significant difference in the number of participants according to gender groups ($t = 2.397$ $p = 0.023$ $p < 0.05$). It was determined that the number of clicking of the males ($\bar{x}=9.69$) was higher than that of female ($\bar{x}=7.03$) (See Table 7.17).

Table 7.17. : *Independent Sample T-Test Results of Mean Net Time, Mean number of Clicks and Mean number of Successful Tasks according to their Gender(* Significance Level at %95 (P < 0.05)).*

	Gender	n	$\bar{x}$	s.d.	Levene Test		t	P
					F	p		
Mean Net Time	Female	16	22.40	7.41	0.085	0.773	0.612	0.545
	Male	16	24.13	8.48				
Mean number of clicks	Female	16	7.03	2.26	2.327	0.138	2.397	0.023*
	Male	16	9.69	3.82				
Mean number of successful tasks	Female	16	6.63	1.15	0.296	0.591	0.946	0.352
	Male	16	7.00	1.10				

H₀: The mean net time, mean number of clicks and mean a number of successful tasks of the participants did not differ significantly according to their gender.

H₁: The mean net time, mean number of clicks and mean a number of successful tasks of the participants did differ significantly according to their gender.

The mean net time of participants was statistically significant according to the self-confidence variation on computer usage ($t = -2.325$ $p = 0.027$ $p < 0.05$). It was determined that the average net times of those who trust in computer usage ($\bar{x} = 26.95$) were higher than those of who stated that their computer usage confidence depends on the given task ($\bar{x} = 20.74$).

Participants' mean number of clicks was found to be statistically significant according to the self –confidence variation on computer usage. ($t = -2.066$ $p = 0.048$ $p < 0.05$). Students who trust their computer usage skills' mean number of clicks ($\bar{x} = 9.78$) was found to be higher than student who stated that their confidence in terms of computer usage differ according to given tasks ($\bar{x} = 7.39$).

According to the results of the t-test to determine whether the participants had significant differences in mean net time, mean number of clicks and mean number of successful tasks compared to self-confidence variation on computer usage, there was no statistically significant difference between the mean number of successful tasks and the self-confidence variation of computer usage ($p > 0.05$) (See Table 7.18).

Table 7.18. : *Independent Sample T-Test Results of Participants' Mean Net Time, Mean number of clicks and Mean number of Successful Tasks according to Self-Confidence variation on computer usage(* Significance Level at %95 ($P < 0.05$)).*

	Self-confidence variation on computer usage	n	$\bar{x}$	s.d.	Levene Test		t	P
					F	p		
Mean Net Time	It depends on the given task.	19	20.74	4.25	14.992	0.001	-2.325	0.027*
	I trust.	13	26.95	10.51				
Mean number of clicks	It depends on the given task.	19	7.39	2.39	2.827	0.103	-2.066	0.048*
	I trust.	13	9.78	4.15				
Mean number of successful tasks	It depends on the given task	19	6.68	1.20	0.205	0.654	-0.778	0.442
	I trust.	13	7.00	1.00				

H₀: Participants' mean net time, the mean number of clicks, and mean a number of successful tasks did not differ significantly from self- confidence variation on computer usage.

H₁: Participants' mean net time, the mean number of clicks, and mean a number of successful tasks did differ significantly from self- confidence variation on computer usage.

In the study, participants' mean net time, mean number of clicks and mean number of successful tasks according to participants' GPA, computer usage times, computer usage skills, daily average computer usage times and frequency of usage of internet web services, age,

frequency of using SIS, the most liked features of the SIS , the most disliked features of the SIS and recommendation to increase usage of SIS, there was no statistically significant relationship between 95% confidence level ($p > 0.05$).

7.4. Discussion and Conclusion

This study was designed to test the research question that is “How do different student groups’ background affect users’ emotion, performance and perceived usability of SIS?”. As far as is known, this study is the first study that examines this research question in the Turkish universities. The results of this study showed that it is possible to use students’ demographic background to predict their preferences for some specific features on SIS.

It was determined that the mean value of EM positive of the IT group was 25.56 ± 8.87 and the mean value of EM negative was 15.50 ± 7.07 . For the OTHER group, EM positive and negative mean scores were found to be 24.31 ± 8.66 and 16.56 ± 7.62 , respectively. Besides, the mean value of SUS-TR of the IT group was 61.72 ± 25.44 and the mean value of SUS-TR of the OTHER group was 61.72 ± 28.82 . Interestingly, the mean value of participants’ SUS-TR scores who were in IT and OTHER group was equal.

As a result of the obtained test result, there exists a significant difference between the perceptions of the participants in OTHER group with EM Negative having different mean ranks of a most liked feature of SIS. According to the Mann Whitney U test results, EM negative levels were found to be significantly different between the participants who have most liked features are ease of use and communication. For instance, the most liked feature was found to be higher in those with ease of use with compared to those who have the favoured feature is communication.

The SUS-TR scores of the OTHER group were found to be statistically significant at 95% confidence level according to the most liked feature in the SIS. According to the Mann

Whitney U test to determine the groups from which the difference originated, it was determined that the SUS scores differed significantly between who prefer accessibility and communication categories. The most liked feature was found to be higher in those with accessibility than those with communication.

It was determined that the difference of the SUS scores of the IT group according to the class groups was statistically significant at the 95% confidence level. According to the Mann Whitney U test to determine the groups from which the difference originated, it was determined that there was a significant difference in scores between first and fourth-grade students. SUS-TR scores of those with fourth grade were lower than those with first grade. This finding suggests that the perceived usability decreases, especially as the duration of use of the system by the IT group increases.

When it comes to performance evaluation results, it has been found that, when students in the IT and students in the OTHER departments are compared to completing the tasks on the system; in the second task, there is a significant difference between the students in the IT and the students in the OTHER Departments. While 94% of students in the IT department completed the task, 81% of the students in the Other departments were able to complete the task. As a result, the students in the IT group were more successful. There is no difference between students in the IT and the students in the OTHER departments in other task assignments. However, participants from both groups performed significantly poorly in task 4 and task 7.

In comparing the net time spent by the tasks with the students in the IT department and the students in the other departments; there is a difference in terms of completion time of the second task between the students in the IT and the OTHER department students. The average time spent by IT students is higher than the spent by OTHER department students.

Also, in general, it is expected that the students in the IT and the related department will have a small number of clicks while completing the task, the results show us that the students in the IT and the related group performed tasks with more clicks than the Other Departments. This is a significant and interesting performance evaluation finding that is noteworthy. These findings can also be explained by the enthusiasm of the IT group students to use information systems' interfaces.

According to Pearson's correlation analysis results for determining the meaningful relationship between the EM positive and EM negative perceptions of the participants and the mean net time, the mean number of clicks and the mean number of successful tasks, no relationship was detected. In addition, according to the results of Pearson correlation analysis to determine the meaningful relationship between the total SUS-TR scores and the mean net time, the number of clicks and the number of successful tasks, there was no significant relationship between the total SUS-TR scores of the participants and the average time, the number of clicks and the number of successful tasks.

According to the results of the independent sample t-test to determine whether the participants had significant differences in terms of mean net time, mean number of clicks and mean number of successful tasks, there was no statistically significant difference between the mean net time and the mean number of successful tasks according to gender groups. However, there was the statistically significant difference in the mean number of click times of participants according to gender groups. It was determined that the number of clicking of the males was higher than that of female.

The mean net time of participants was statistically significant according to the self-confidence variation on computer usage. It was determined that the average net times of those who trust in computer usage were higher than those of who stated that their computer usage

confidence depends on the given task. As self-confidence increases, tasks are expected to be completed with fewer net time. However the findings are exactly the opposite.

Participants' mean number of clicks was found to be statistically significant according to the self –confidence variation on computer usage. Students who trust their computer usage skills' mean number of clicks was found to be higher than student who stated that their confidence in terms of computer usage differ according to given tasks. As self-confidence increases, tasks are expected to be completed with fewer clicks but the findings are in a kind of interesting way the opposite. These findings can also be explained by the enthusiasm of the self-confident group students to use information systems' interfaces.

In the study, participants' mean net time, mean number of clicks and mean number of successful tasks according to participants' GPA, computer usage times, computer usage skills, daily average computer usage times and frequency of usage of internet web services, age, frequency of using SIS, the most liked features of the SIS , the most disliked features of the SIS and recommendation to increase usage of SIS, there was no statistically significant relationship between 95% confidence level.

Therefore, analyzing users' emotional variations amongst different student groups is valuable because focusing on these variations has the potential to raise user experiences of users. Besides, developers will be attentive to suitable features to use on websites if they are aware of their visitors' demographic background. For these reasons, focusing on students' background has a potential to raise their satisfaction with the system.

8. DISCUSSION AND CONCLUSIONS

The first subsection of this final chapter demonstrates the study results found and emphasizes the main contributions of the programme of study (See section 8.1). Next, the limitations of the research are discussed (See section 8.2). Then, future work is explained by identifying further research perspectives (See section 8.3). The thesis concludes with implications (See section 8.4) and guidelines (See section 8.5) that are reviewed based on the study results obtained.

8.1. Synthesis of Main Results from this Programme of Research

The thesis introduces a detailed literature review, related to HCI, usability, usability evaluation methods, user emotion and experience in Chapter 2. These reviews are worthwhile to researchers who study and work on these topics.

In Chapter 3, the Turkish version of the Emotion Word Prompt List (EWPL-TR) is introduced, this is a Turkish version of the EWPL. With permission from the EWPL's main author Professor Dr. Helen Petrie, the EWP-TR was developed by using professional interpretation techniques. Also, after the item 15 was subtracted, it was evaluated for its reliability and construct validity, and was found to have good reliability and validity. Therefore, now it is usable with Turkish learners. Turkish researchers who wish to undertake research with Turkish participants related to the EWPL can utilize this questionnaire.

In Chapter 4, the Turkish version of the System Usability Scale (SUS-TR) is introduced, this is a Turkish version of the SUS. With permission from the SUS's main author Dr. John Brooke, the SUS-TR was developed by using professional translation techniques. Also, it was evaluated for its reliability and construct validity, and was found to have good reliability and validity. Therefore, now it is usable with Turkish learners. Turkish researchers who wish to carry out research with Turkish participants related to SUS-TR can use this survey.

In Chapter 5, the study was designed to test the research question that is “How do different student groups’ background affect users’ emotion and perceived usability of SIS?” For this purpose EWPL-TR and SUS-TR applied 324 undergraduate students. The study showed that different student demographic backgrounds affect users’ emotion but do not affect perception of the usability apart from age and most disliked features of the SIS questions.

In Chapter 6, the pilot experiment has been carried out. Pilot study’s aim was before experimenting large student groups to identify possible problems. So that, this study showed that this study’s method can be used to evaluate students’ emotions, performance and perceived usability on a specific SIS.

In Chapter 7, the research question (How do different students groups’ demographic background affect users’ experience, performance and perceived usability of SIS has been addressed. This study showed that different student some demographic backgrounds affect users’ emotion, performance and perception of the usability of the SIS. Since our sample group is very small, larger groups of students should be studied in order to have more meaningful results.

8.2. Limitations of the studies

There were some critical methodological impediments in this examination. A confinement of this program of research is the generally limited student samples who took an interest in the three principle contemplates. Members of the examinations displayed in sections were students, all from same university in the Turkey. Likewise, this investigation incorporates an usability test with a predetermined number of students in a solitary foundation as they finished a set number of academic assignments. Likewise students who are acquainted with the framework may tweak his or her interface to a shifting degree, which isn't

reflected in this examination and may have impacted the outcomes. In addition, this study also did not evaluate the usability of student information systems in terms of academic and administrative staff. Besides, eye tracking data may give more accurate clues about usability of the tested system.

8.3. Future work

In order to overcome some of the limitations of the programme of research and generalize the outcomes from the research, further study needs to be done with a wider range of users and using different SIS. It would be interesting to conduct the experiments in more realistic settings, with users using their systems in their own home or work environments.

Keeping in mind the end goal to defeat a some of the restrictions of the program of research and sum up the results from the exploration, additionally the next study should be conducted with a more extensive scope of students and utilizing diverse SIS. It is interesting to direct the investigations in more realistic settings, with students utilizing their frameworks in their own home or workplaces.

An important next step will be to provide further validation of the EWPL-TR and SUS-TR by using it with larger and more varied samples of Turkish people. Also usability of the SIS can be evaluated different groups of users such as academic and administrative staff. In addition, the scales used in the research, can be analysed by structural equation modeling which is as a second generation data analysis technique, provides a systematic and comprehensive study of a complex research problem in a single process, by modeling the relationships between many dependent and independent variables, as compared to first generation statistical techniques such as regression.

8.4. Implications for SIS developers

In light of consequences of this program of research, plainly system engineers ought to know about the qualities of their intended interest groups, for example, unique student groups' backgrounds.

Moreover, perceived usability of systems can be enhanced by dissecting clients' emotional varieties among various demographic gatherings. When creating systems, web designers should remember that they are designing to clients with various student groups.



REFERENCES

- Bahrack, H. P., Hall, L. K., Goggin, J. P., Bahrack, L. E., & Berger, S. A. (1994). Fifty years of language maintenance and language dominance in bilingual Hispanic immigrants. *Journal of Experimental Psychology: General*, 123, 264–283.
- Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*, 24(6), 574–594. <https://doi.org/10.1080/10447310802205776>
- Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, 4(3), 114–123. <https://doi.org/66.39.39.113>
- Barnum, M. C. (2011). Praise for Usability Testing Essentials. *Usability Testing Essentials*. <https://doi.org/http://dx.doi.org/10.1016/B978-0-12-375092-1.00023-4>
- Battleson, B., Booth, A., & Weintrop, J. (2001). Usability testing of an academic library Web site: a case study. *The Journal of Academic Librarianship*, 27(3), 188–198. [https://doi.org/10.1016/S0099-1333\(01\)00180-X](https://doi.org/10.1016/S0099-1333(01)00180-X)
- Blažica, B., & Lewis, J. R. (2015). A Slovene translation of the system usability scale: The SUS-SI. *International Journal of Human-Computer Interaction*, 31(2), 112–117. <https://doi.org/10.1080/10447318.2014.986634>
- Blythe, M. and Monk, A. (2002). Notes towards an ethnography of domestic technology. In *Proceedings of DIS 2002*, pp. 276-281.
- Brooke, J. (1996). SUS - A quick and dirty usability scale. *Usability Evaluation in Industry*, 189(194), 4–7. <https://doi.org/10.1002/hbm.20701>
- Brooke, J. (2013). SUS : A retrospective. *Journal of Usability Studies*, 8(2), 29–40. <https://doi.org/10.1074/jbc.R115.675280>
- Büyüköztürk, Ş. (2002). Faktor Analizi: Temel kavramlar ve Ölçek Gelistirmede Kullanımı. *Kuram ve Uygulamada Eğitim Yönetimi Dergisi*, (32), 470–483.
- Büyüköztürk, Ş. (2010). *Sosyal Bilimler için Veri Analizi El Kitabı*, Pegem Akademi Yayıncılık, Ankara.
- Büyüköztürk, Ş., Çakmak, E. K., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2017). Bilimsel araştırma yöntemleri. *Pegem Atıf İndeksi*, 1-360.
- Carcary, M., Long, G., & Remenyi, D. (2007). The Implementation of a New Student Management Information System (MIS) at an Irish Institute of Technology – An Ex Post Evaluation of its Success. *Electronic Journal of Information Systems Evaluation*, 10(1), 31–44. Retrieved from www.ejise.com
- Chin, J. P., Diehl, V. a, & Norman, L. K. (1988). Development of an instrument measuring

- user satisfaction of the human-computer interface. CHI '88- Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 213–218.
<https://doi.org/10.1145/57167.57203>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests* If~ j. cronbach, 16(3), 297–298.
- Çokluk, Ö., Şekercioğlu, G. ve Büyüköztürk, Ş. (2012). Sosyal bilimler için çok değişkenli istatistik: SPSS ve lisrel uygulamaları, Pegem Akademi Yayıncılık, Ankara.
- Davis, D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13, 319–339.
- Delgado, P., Guerrero, G., Goggin, J. P., & Ellis, B. B. (1999). Self-assessment of linguistic skills by bilingual Hispanics. Hispanic Journal of Behavioral Sciences, 21, 31–46.
- Dianat, I., Ghanbari, Z., & AsghariJafarabadi, M. (2014). Psychometric properties of the persian language version of the system usability scale. Health Promotion Perspectives, 4(1), 82–9. <https://doi.org/10.5681/hpp.2014.011>
- Diaper, D. ; S. A. N. (2004). The Handbook of Task Analysis for Human Computer Interaction. The handbook of task analysis for humancomputer interaction. Lawrence Erlbaum Associates, Inc., Publishers 10 Industrial Avenue Mahwah, New Jersey 07430: LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS.
- Dillon, A. (2001) Usability evaluation. In W. Karwowski (ed.) Encyclopedia of Human Factors and Ergonomics, London: Taylor and Francis.
- Dix, A. U., Finlay, J. M. U., Abowd, G. I. of T., & Beale, R. of B. (2004). Human - Computer Interaction. Pearson.
- Dumas, S. Jaseph, R. J. (1999). A Practical Guide to Usability Testing.
- Faulkner, L. (2003). Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. Behavior Research Methods, Instruments, & Computers, 35(3), 379–383. <https://doi.org/10.3758/BF03195514>
- Forlizzi, J., & Battarbee, K. (2004). Understanding Experience in Interactive Systems. Human-Computer Interaction Institute, 37(2), 261–268.
<https://doi.org/10.1145/1013115.1013152>
- Frøkjær, E., Hertzum, M., & Hornbæk, K. (2000). Measuring Usability : Are Effectiveness , Efficiency , and Satisfaction Really Correlated ? ACM CHI 2000 Conference on Human Factors in Computing Systems, 2(1), 345–352.
<https://doi.org/10.1145/332040.332455>
- Garnefski, N., & Kraaij, V. (2007). The cognitive emotion regulation questionnaire: Psychometric features and prospective relationships with depression and anxiety in adults. European Journal of Psychological Assessment, 23(3), 141–149.
<https://doi.org/10.1027/1015-5759.23.3.141>

- Gemmell, M., & Pagano, R. (2003). A Post-Implementation Evaluation of a Student Information System in the UK Higher Education Sector. *Electronic Journal of Information Systems Evaluation*, 6(2), 95–106.
- Gould, J. D., & Lewis, C. (1985). Designing for usability: key principles and what designers think. *Communications of the ACM*, 28(3), 300–311.
<https://doi.org/10.1145/3166.3170>
- Hackos, J., & Redish, J. (1998). User and Task analysis for Interface design.
<https://doi.org/10.1109/TPC.1999.784573>
- Harniss, M. K., Epstein, M. H., Ryser, G., & Pearson, N. (1999). The Behavioral and Emotional Rating Scale: Convergent validity. *Journal of Psychoeducational Assessment*, 17(1), 4–14. Retrieved from <http://proxy-remote.galib.uga.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1999-13280-001&site=ehost-live>
- Harper, B. D., & Norman, K. L. (1998). Improving user satisfaction: The questionnaire for user interaction satisfaction version 5.5. *Proceedings of Mid Atlantic Human Factors Conference*, 224–228.
- Harrison, C. (2009). Exploring emotional web experience: more than just usability and good design. PhD Thesis, University of York, UK.
- Head AJ (1999) Design wise. Thomas H Hogan Sr, Medford
- Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human Computer Studies*, 64(2), 79–102.
<https://doi.org/10.1016/j.ijhcs.2005.06.002>
- IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.
- Isemonger, I. & Sheppard, C. (2007). A construct-related validity study on a Korean version of the perceptual learning styles preference questionnaire. *Educational and Psychological Measurement*, 67(2), 357-368
- ISO, 1998. Ergonomic requirements for office work with visual display terminals (VDTs)-Part 11: guidance on usability—Part 11: guidance on usability (ISO 9241-11:1998).
- Jordan, P. (1998). An introduction to Usability. London: Taylor & Francis Ltd.
- Jöreskog, K.G. & Sörbom, D. (2015). LISREL 9.20 for Windows [Computer software]. Skokie, IL: Scientific Software International, Inc.
- Kadirhan, Z., Gül, A., & Battal, A. (2015). SİSTEM KULLANILABİLİRLİK ÖLÇEĞİ: GEÇERLİK VE GÜVENİRLİK ÇALIŞMASI.
- Kampf, M., & Payne, S. (2000). What is beautiful is usable. *Interacting with Computers*,

13(December), 127–145.

- Karagöz, Y. ve Kösterelioğlu, İ. (2008). “İletişim becerileri değerlendirme ölçeğinin faktör analizi metodu ile geliştirilmesi” *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 21, 81-98.
- Kim, J., & Moon, J. Y. (1998). Designing towards emotional usability in customer interfaces—trustworthiness of cyber-banking system interfaces. *Interacting with Computers*, 10(97), 1–29. [https://doi.org/10.1016/S0953-5438\(97\)00037-4](https://doi.org/10.1016/S0953-5438(97)00037-4)
- Kuhn, K. (2000). Problems and benefits of requirements gathering with focus groups: A case study. *International Journal of Human-Computer Interaction*, 12(3), 309–325. https://doi.org/10.1207/S15327590IJHC1203&4_4
- Kujala, S., & Miron-Shatz, T. (2013). Emotions, experiences and usability in real-life mobile phone use. *Proceedings of CHI 2013*, (November), 1061–1070. <https://doi.org/10.1145/2470654.2466135>
- Laudon, K. C., & Laudon, J. P. (1968). *Management Information Systems Managing the Digital Firm*. Management Decision (Vol. 2). <https://doi.org/10.1108/eb000831>
- Lewis, C., 1982. Using the “thinking-aloud” method in cognitive interface design. Research Report RC9265, IBM T.J. Watson Research Center.
- Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires : Psychometric evaluation and instructions for use IBM Computer Usability Satisfaction Questionnaires : Psychometric Evaluation and Instructions for Use. *International Journal of Human- Computer Interaction*, 7(1), 57–78. <https://doi.org/10.1037/t32698-000>
- M.Sweeney, M.Maguire, B. S. (1993). Evaluating user-computer interaction: a framework. *International Journal of Man-Machine Studies*, 38(4).
- Mack, Z., & Sharples, S. (2009). The importance of usability in product choice: A mobile phone case study. *Ergonomics*, 52(12), 1514–1528. <https://doi.org/10.1080/00140130903197446>
- Macleod, M (1994) Usability in context: Improving quality of use. In: G Bradley and HW Hendricks (eds.) *Human Factors in Organizational Design and Management - IV - Proceedings of the International Ergonomics Association 4th International Symposium on Human Factors in Organizational Design and Management*, (Stockholm, Sweden, May 29 - June 1 1994). Amsterdam, Elsevier.
- Martins, A. I., Rosa, A. F., Queirós, A., Silva, A., & Rocha, N. P. (2015). European portuguese validation of the system usability scale (SUS). *Procedia Computer Science*, 67(Dsai), 293–300. <https://doi.org/10.1016/j.procs.2015.09.273>
- Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion*, 23(2), 209–237. <https://doi.org/10.1080/02699930802204677>

- Maxwell, B. (1996). Translation and cultural adaptation of the survey instruments. In: Martin, M. O. & Kelly, D. L. (eds.). Third International Mathematics and Science Study (TIMSS) Technical Report, Volume I: Design and Development. Chestnut Hill, MA: Boston College.
- Molich, R., Nielsen, J., 1990. Improving a human–computer dialogue. *Communications of the ACM* 33 (3), 338–348
- Mukerjee, S. (2012). Student information systems – implementation challenges and the road ahead. *Journal of Higher Education Policy and Management*, 34(1), 51–60.
<https://doi.org/10.1080/1360080X.2012.642332>
- Nielsen, J., & Molich, R. (1990). Heuristic Evaluation of user interfaces. CHI '90 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, (April), 249–256. <https://doi.org/10.1145/97243.97281>
- Nielsen, J. (1992). Finding usability problems through heuristic evaluation. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '92, 373–380.
<https://doi.org/10.1145/142750.142834>
- Nielsen, J. (1993). Usability Engineering. Morgan Kaufmann Pietquin O and Beaufort R, 44(1/2002), 362. <https://doi.org/10.1145/1508044.1508050>
- Nielsen, J., & Levy, J. (1994). Measuring usability: preference vs. performance. *Communications of the ACM*. <https://doi.org/10.1145/175276.175282>
- Nielsen, J. (01.01.1995). Usability 101: Introduction to usability [Available online at: <https://www.nngroup.com/articles/ten-usability-heuristics/>], Retrieved from August 22th, 2018.
- Nielsen, J., & Pernice, K. (2010). Eyetracking web usability. New Riders.
- Nielsen, J. (04.01.2018). Usability 101: Introduction to usability [Available online at: <https://www.nngroup.com/articles/usability-101-introduction-to-usability/>], Retrieved from February 22th, 2018.
- Nordaliela, M. R., Suriani, H., & Nathaniel, E. L. (2013). Usability Analysis of Students Information System in a Public University. *Usability Analysis of Students Information System in a Public University*, 4(6), 806–810.
- Nunnally, J. C. (1978). Psychometric theory (2nd ed.). New York: McGraw-Hill.
- Barnette, J. J. (2000). Effects of stem and Likert response option reversals on survey internal consistency: If you feel the need, there is a better alternative to using those negatively worded stems. *Educational and Psychological Measurement*, 60, 361–370.
- Peter, C., & Beale, R. (2008). Affect and Emotion in Human Computer Interaction, From Theory to Applications. (C. Peter & R. Beale, Eds.), Affect and Emotion in HCI, LNCS. https://doi.org/10.1007/978-3-540-85099-1_10

- Petrie, H., & Harrison, C. (2009). Measuring users' emotional reactions to websites. *Proceedings of the 27th International Conference Extended Abstracts on Human Factors in Computing Systems CHI EA 09*, 3847. <https://doi.org/10.1145/1520340.1520582>
- Petrie, H., & Precious, J. (2010). Measuring user experience of websites: think aloud protocols and an emotion word prompt list. *CHI'10 Extended Abstracts on Human Factors in ...*, 3673–3678. <https://doi.org/10.1145/1753846.1754037>
- Pilotte, W. J., & Gable, R. K. (1990). The impact of positive and negative item stems on the validity of a computer anxiety scale. *Educational and Psychological Measurement*, 50, 603–610.
- Rubin, J., & Chisnel, D. (2014). *Handbook of Usability Testing*. Igarss 2014. <https://doi.org/10.1007/s13398-014-0173-7.2>
- Rubin, J., & Chisnell, D. (2008). *Handbook of usability testing [electronic resource] : How to plan, design, and conduct effective tests (2nd ed.)*. Indianapolis, IN: Wiley Pub. <https://doi.org/10.1007/s13398-014-0173-7.2>
- Sauro, J., & Lewis, J. R. (2011). When designing usability questionnaires, does it hurt to be positive? In *proceedings of CHI 2011* (pp. 2215–2223). Vancouver, Canada: ACM.
- Sauro, J., & Lewis, J. R. (2012). *Quantifying The User Experience. Practical Statistics for user research*. Morgan Kaufmann. <https://doi.org/10.1007/s13398-014-0173-7.2>
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research Online*, 8(2), 23–74. <https://doi.org/10.1002/0470010940>
- Schmitt, N., & Stuits, D. (1985) Factors defined by negatively keyed items: The result of careless respondents? *Applied Psychological Measurement*, 9, 367–373.
- Scholtz, J. (2003). *Usability Evaluation User-Centered Evaluations*. Institute of Standards and Technology, 379(1478–1–PB), 1–8.
- Schriesheim, C. A., & Hill, K. D. (1981). Controlling acquiescence response bias by item reversals: The effect on questionnaire validity. *Educational and Psychological Measurement*, 41, 1101–1114.
- Sears, A., & Jacko, J. (2008). *The Human – Computer Interaction Handbook*. 2008 by Taylor & Francis Group, LLC.
- Seneler, C. (2014) *The impact of learning styles and cultural background on users' experience of websites*. PhD thesis, University of York.
- Sharp, H., Rogers, Y., & Preece, J. (2007). *Interaction design: beyond human-computer interaction*. Book, 11, 773. <https://doi.org/10.1162/leon.2005.38.5.401>
- Shneiderman, B., & Plaisant, C. (2005). *Designing the user interface: strategies for effective*

- human-computer interaction. British dental journal (Vol. 215).
<https://doi.org/10.1038/sj.bdj.2013.932>
- Smith, S. L., & Mosier, J. N. (1986). Guidelines for Designing User Interface Software, (August), 329. <https://doi.org/10.1001/archdermatol.2011.235>
- Şengel, E. (2014). Discovering how students search a university web site: A comparative usability case study for PC and mobile devices. Turkish Online Journal of Educational Technology.
- Tabachnick, B. G. ve Fidell, L. S. (2007). Using multivariate statistics. Boston: Pearson Education Inc.
- Tabachnick and Fidell, 2013 B.G. Tabachnick, L.S. Fidell Using Multivariate Statistics (sixth ed.) Pearson, Boston (2013)
- Task Analysis [Available online at: <https://www.usability.gov/how-to-and-tools/methods/task-analysis.html>], Retrieved from: 25th August 2017
- Thompson, B., & Levitov, J. E. (1985). Using microcomputers to score and evaluate test items. Collegiate Microcomputer, 3, 163-168.
- Thüring, M., & Mahlke, S. (2007). Usability, aesthetics and emotions in human-technology interaction. International Journal of Psychology, 42(4), 253–264.
<https://doi.org/10.1080/00207590701396674>
- Tullis, T. S., & Stetson, J. N. (2004). A comparison of questionnaires for assessing website usability.
- Tullis, T., & Albert, B. (2013). Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics: Second Edition. <https://doi.org/10.1016/C2011-0-00016-9>
- Usability Evaluation Basics [Available online at: <https://www.usability.gov/what-and-why/usability-evaluation.html>], Retrieved from 24th August 2017
- Vladimir Zwass (n.d). *Information Systems* [Available online at: <https://www.britannica.com/topic/information-system>], Retrieved from: 15th May 2017
- Wharton, C., Rieman, J., Lewis, C., Polson, P., 1994. The cognitive walkthrough method: a practitioner's guide. In: Nielsen, J., Mack, R.L. (Eds.), Usability Inspection Methods. Wiley, New York, pp. 105–140.
- Wong, N., Rindfleisch, A., & Burroughs, J. (2003). Do reverse-worded items confound measures in cross-cultural consumer research? The case of the material values scale. Journal of Consumer Research, 30, 72–91.

APPENDIX A: List of Abbreviation

ACM: Association for computing machinery
AGFI: Adjusted Good Fit Index
CFA: Confirmatory Factor Analysis
CFI: Comparative Fit Index
CHI: Computer and human interaction
CSUQ: Computer System Usability Questionnaire
Df: Degrees of freedom
EFA: Exploratory Factor Analysis
EWPL: Emotion Word Prompt List
EWPL-TR: Turkish Version of the Emotion Word Prompt List
GFI: Goodness of Fit Index
GPA: Grade Point Average
HCI: Human – computer interaction
HMI: Human – machine interaction
IS: Information Systems
ISO: International organization for standardization
IT: Information technology
KMO: Kaiser-Meyer-Olkin
MMI: Man – machine interaction
NFI: Normed Fit Index
OAQ: Overall assessment questions
QUIS: Questionnaire for User Interface Satisfaction
RMSEA: Root Mean Square Error of Approximation
Sd: Standard Deviation
SIS: Student Information Systems
SUS: System Usability Scale
SUS-TR: Turkish Version of the System Usability Scale
UX: User Experience

APPENDIX B

Table B.1. : *Original version of the EWPL*

Word List	1 = Never	2	3	4	5 = Very much
Amused					
Annoyed					
Bored					
Confident					
Confused					
Creative					
Curious					
Disappointed					
Frustrated					
Happy					
Interested					
Hopeful					
Pleased					
Relieved					
Surprised					
Unsure					

Table B.2. : *Finalized Turkish version of the EWPL (EWPL-TR)*

Kelime Listesi	1 = Hiç	2	3	4	5 = Çok
Mutlu					
Sinirlenmiş					
Sıkılmış					
Kendinden Emin					
Kafası Karışmış					
Yaratıcı					
Meraklı					
Hayal Kırıklığına Uğramış					
Usanmış					
Mutlu					
İlgili					
Umutlu					
Hoşnut					
Rahatlamış					
Güvensiz					

Table B.3.: *Original version of the SUS*

System Usability Scale (SUS)	1 = Strongly disagree	2	3	4	5 = Strongly agree
1. I think that I would like to use this system frequently.					
2. I found the system unnecessarily complex.					
3. I thought the system was easy to use.					
4. I think that I would need the support of a technical person to be able to use this system.					
5. I found the various functions in this system were well integrated.					
6. I thought there was too much inconsistency in this system.					
7. I would imagine that most people would learn to use this system very quickly.					
8. I found the system very cumbersome/awkward to use.					
9. I felt very confident using the system.					
10. I needed to learn a lot of things before I could get going with this system.					

Table B.4. : Turkish version of the System Usability Scale (The SUS-TR)

Sistem Kullanılabilirlik Ölçeği	1 = Hiç katılmıyorum	2	3	4	5 = Tamamen katılıyorum
1. Bu sistemi sıklıkla kullanmak isteyeceğimi düşünüyorum.					
2. Bu sistemi gereksiz bir şekilde karmaşık buldum.					
3. Bu sistemin kullanımının kolay olduğunu düşündüm.					
4. Bu sistemi kullanabilmek için daha teknik bir kişinin desteğine ihtiyaç duyacağımı düşünüyorum.					
5. Bu sistemdeki çeşitli fonksiyonları iyi entegre edilmiş buldum.					
6. Bu sistemde çok fazla tutarsızlık olduğunu düşündüm.					
7. Birçok insanın bu sistemi kullanmayı çok çabuk öğreneceğini sanıyorum.					
8. Bu sistemin kullanımını çok elverişsiz buldum.					
9. Bu sistemi kullanırken kendimden çok emin hissettim.					
10. Bu sistemde bir şeyler yapabilmek için öncelikle bir çok şey öğrenmem gerekti.					

APPENDIX C

C.1. Öğrenci Bilgi Sistemi Görevler Listesi

Aşağıda Öğrenci Bilgi Sisteminin kullanımına ilişkin 8 adet görev tanımlanmaktadır. Lütfen belirtilen görevleri sırasıyla gerçekleştiriniz ve her görev için Öğrenci Bilgi Sistemi web sitesinin ana sayfasından başlayınız. Herhangi bir görevi gerçekleştirmeden önce bu görevi sesli olarak okuyunuz ve anlaşılmayan bir unsurun olması halinde araştırmacıya danışınız.

Table C.1. : Öğrenci Bilgi Sistemi Görevler Listesi

	GÖREVLER
1	Kullanıcı adınız ve Şifreniz ile sisteme giriş yapınız.
2	Transcriptinize ulaşmak istiyorsunuz.Bu bilgiye hangi sayfadan ulaşabileceğinizi bulunuz.
3	Akademik Danışmanınıza Öğrenci Bilgi Sistemi üzerinden mesaj göndermek istiyorsunuz.Bu işlemi gerçekleştirmek için kullanmanız gereken sayfayı bulunuz.
4	Öğrenci Bilgi Sistemi yardım videolarına ulaşmak istiyorsunuz.Bu durumda nasıl bir yol izlemeniz gerektiğini bulunuz.
5	Sisteme giriş şifrenizi değiştirmek istiyorsunuz.Bu durumda nasıl bir yol izlemeniz gerektiğini bulunuz.
6	Bölümünüzden ders seçimi yapmak istiyorsunuz.Bu işlemi gerçekleştireceğiniz ekranı bulunuz
7	Bölümünüz dışı ders seçimi yapmak istiyorsunuz.Bu işlemi gerçekleştirebileceğiniz ekranı bulunuz.
8	Mali Onay Bitiş Tarihi'ne ulaşmak istiyorsunuz. Bu durumda nasıl bir yol izlemeniz gerektiğini bulunuz.

C.2. Gönüllü Katılım Formu

Bu çalışmanın temel amacı Öğrenci Bilgi Sisteminin değişik öğrenci grupları arasındaki kullanıcı duygularını, algılanan kullanılabilirliği ve performanslarını değerlendirmektir. Bu testin amacı kesinlikle sizin bilginizi test etmek değil, yalnızca Öğrenci Bilgi Sisteminin kullanılabilirliğinin değerlendirmesini yapmaktır.

Test Aşamaları:

Öncelikle test sırasında, sistem kullanımına yönelik hazırlanan ve toplam 8 sorudan oluşan “görevleri” siteyi kullanarak gerçekleştirmeniz istenecektir.

Görevler listesi size yazılı olarak verilecektir. Her bir görevi gerçekleştirmeden önce sesli olarak okumanızı ve anlaşılmayan noktalar var ise test gözlemcisine sormanızı rica ediyoruz.

Test gözlemcisi test sırasında yanınızda bulunarak, test ile ilgili notlar alacaktır.

Her yeni “göreve” sistemin ana ekranından başlamanız gerekmektedir.

Görevi tamamladığınızda, ya da başka bir göreve geçmek istediğinizde lütfen bunu gözlemciye sözlü olarak bildiriniz.

Toplam test süresinin yaklaşık olarak 30 dakika süreceği tahmin edilmektedir. Ancak istediğiniz zaman testi bırakmakta serbestsiniz.

Test gözlemcisi site kullanımıyla ilgili olarak yönlendirme yapmamak koşuluyla sorularınızı cevaplayacaktır.

Verilerinizin analizini kolaylaştırmak amacıyla , test sırasındaki ekran görüntüleri ve ses kaydı kaydedilecektir.

Görev aşaması tamamlandıktan sonra sistem hakkında duygularınızı ölçmek amacıyla “Duygusal Kelime İstem Listesi” anketi uygulanacaktır.

Daha sonra sistem kullanımının ne seviyede olduğunu belirlemek amacıyla “Sistem Kullanılabilirlik Ölçeği” uygulanacaktır.

Hemen ardından sistem hakkındaki görüşlerinizi almak amacıyla “Öğrenci Bilgi Sistemi, Genel Değerlendirme Anketi” size yöneltilecektir.

Son olarak da demografik bilgilerinizi almak amacıyla “Kullanıcı Bilgi Anketi”ni yanıtlamanız istenecektir.

Kişisel Haklar ve Bilgilerin Gizliliği

Bu form ile, çalışma sonucu elde edilen verilerin akademik amaçlı kullanılabilmesi için izniniz istenmektedir. Test için toplanan veriler hiçbir koşulda adınız ile ilişkilendirilmeyecektir. Geribildirim için denizhan.demirkol@adu.edu.tr adresi ile irtibat kurabilirsiniz.

Test başlamadan önce sormak istediğiniz sorular var ise lütfen şimdi sorunuz !

Test öncesinde gözlemci tüm sorularımı yanıtladı. Teste katılmak istiyorum.

Katılımcının imzası:

Test Tarihi ve Saat:

C.3. Öğrenci Bilgi Sistemi Genel Değerlendirme Anketi

1) Öğrenci Bilgi Sisteminde en beğendiğiniz özellik nedir? Neden?

2) Öğrenci Bilgi Sisteminde en beğenmediğiniz özellik nedir? Neden?

3) Öğrenci Bilgi Sisteminin kullanılabilirliğini arttırmak için ne yapılmasını önerirsiniz? Neden?

C.4. Kullanıcı Bilgi Anketi

1.Cinsiyet:

Kadın ☐ Erkek ☐

2. Yaşınız?

.....

3. Eğitim düzeyi:

Lisans ☐

Yüksek Lisans ☐

Doktora ☐

4.Üniversitede kaçınıcı seneniz?

☐1

☐2

☐3

☐4

☐5

☐6+

5. Kayıtlı olduğunuz bölüm nedir? Anadal ve varsa yandal ?

6.Bulunduğunuz dönem itibariyle genel akademik not ortalamanız kaçtır?

3.50 ile 4.00 arasında ☐

3.00 ile 3.49 arasında ☐

2.50 ile 2.99 arasında ☐

2.00 ile 2.49 arasında ☐

2.00'nin altında ☐

Bilmiyorum ☐

7. Kaç yıldır bilgisayar kullanıyorsunuz?

3-5 yıl arası kullanıyorum. ☐

5-7 yıl arası kullanıyorum. ☐

7 – 10 yıl arası kullanıyorum. ☐

10 yıl ve üzeri kullanıyorum. ☐

8.Bilgisayar kullanma becerinizi nasıl değerlendirirsiniz?

Başlangıç seviyesinde ☐

Orta seviye ☐

Orta- İleri seviye ☐

İleri seviye ☐

9.Bilgisayar kullanımı konusunda kendinize ne kadar güvenirsiniz?

Hiç Güvenmiyorum. ☐

Genellikle yardıma ihtiyaç duyarım. ☐

Verilen göreve bağlıdır. ☐

Güveniyorum. ☐

10. Günde ortalama kaç saat bilgisayar kullanıyorsunuz?

11.İnternet – Web hizmetlerini kullanım sıklığı:

Hergün kullanıyorum.☐

Haftada birkaç gün kullanıyorum.☐

Ayda birkaç gün kullanıyorum.☐

12. Öğrenci Bilgi Sistemi kullanım sıklığı:

Hergün kullanıyorum.☐

Haftada birkaç gün kullanıyorum.☐

Ayda birkaç gün kullanıyorum.☐

Yılda birkaç gün kullanıyorum.☐

APPENDIX D

D.1 GLOSSARY

Back-translation technique: A back-translation technique is a translation of an instrument that has been already translated into a target language back into the source language (Maxwell, 1996).

Construct validity: Construct validity refers to a degree to which a test measures what it claims to be measuring.

Content analysis: Content analysis is a technique for making inferences by methodically and objectively identifying particular characteristics of texts (Holsti, 1968, p. 608).

Cronbach alpha: Cronbach alpha can be defined as a measurement of internal consistency.

Eye tracking: Eye tracking equipment allows researchers to see exactly what users see and where they look by tracking their eye movements and the length of time they fixate on a certain part of a screen or a word, phrase, or object on the screen (Barnum, 2011).

Factor analysis: The use of one of several methods for reducing a set of variables to a lesser number of new variables.

Human Computer Interaction: a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them.

Item Analysis: Item analysis uses statistics and expert judgment to evaluate tests based on the quality of individual items, item sets, and entire sets of items, as well as the relationship of each item to other items. It explores the performance of items considered separately either in relation to some external criterion or in relation to the remaining items on the test (Thompson & Levitov, 1985).

Multiple forward translation technique: A multiple forward translation technique is the translation of an instrument from the source language into the target language independently by a number of translators (Maxwell, 1996).

Reliability analysis : Reliability was assessed using coefficient alpha (Cronbach, 1951)

Usability: ISO standard 9,241-11 Guidance on Usability (1998) defines usability as ‘the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use’.

Varimax Rotation: It is used to simplify the expression of a particular sub-space in terms of just a few major items each.

APPENDIX E

E.1. Curriculum Vitae

DENİZHAN DEMİRKOL

E-mail : denizhan.demirkol@adu.edu.tr , denizhandemirkol@gmail.com

Date of Birth : 16.09.1990

Place of Birth : AYDIN

Nationality : TURKEY

EDUCATION :

2016- 2018 June **Yeditepe University (Master Degree)**
Institute of Social Sciences
Management Information Systems
GPA: 3.56/4

THESIS TITLE : EVALUATION OF STUDENT INFORMATION SYSTEM (SIS)
IN TERMS OF USER EMOTIONS,PERFORMANCE AND PERCEIVED USABILITY : A
TURKISH UNIVERSITY CASE

2009- 2013 **Yeditepe University (Bachelor of Science)**
Information Systems and Technologies
Graduation Date :12.06.2013
Grade-Point Average: **78.30/100**

2008 – 2009 **Yeditepe University**
English Preparation Class

2004 – 2008 **Süleyman Demirel Anatolian High School (AYDIN)**
Weighted High School GPA: 89,04/100

Scholarly Interests & Research Areas

Human-Computer Interaction, Design Tools, User Interface Engineering, Personalization, Adaptive Systems, User-Modeling, Educational Systems, Learner Modeling, Learning Styles, Data Mining, Machine Learning, Medical Informatics

Languages : **English - Advanced**
2017 Turkey Council of Higher Education (CoHE) Language

Grade: **91,250 / 100**
German - Beginner

Work Experiences**Adnan Menderes University – Soke Faculty of Business****Research Asisstant****January 2016 –**

IBTECH International Informatics and Communication Technologies
 Alternative Distribution Channels Web Intranet Business Analyst
 August 2013- January 2014

IBTECH International Informatics and Communication Technologies
 Alternative Distribution Channels Web Intranet Business Analyst
 05.08.2012-10.09.2012

GARANTİ TECHNOLOGY / Information Technologies and Trade Incorporated Company
 Networking Trainee
 25.06.2012-03.08.2012

International Proceedings:

Demirkol,D , Genc F.N. (2018). Legal Status of Bitcoin: Examples with Current Assessment from Turkey and the World, presented at 1st International Congress of Political, Economic and Financial Analysis- 2018(PEFA'2018), Nazilli (Aydın), Turkey on April 27

Demirkol,D , Şeneler Ç. (2018). Evaluation of a student information system (SIS) in terms of user emotions, performance and perceived usability: A pilot study, presented at XV. International European Conference on Social and Behavioral Sciences, Kuşadası (Davutlar), Aydın, Turkey on February 1 2018: E-BWN Publishing.

Erdebilli B. ,Demirkol D. ,AHP Approach for Selecting Enterprise Resource Planning (ERP),6th International Conference on New Challenges in Management and Business, Dubai, 11.02.2017

ATTENDENCE AND CERTIFICATIONS:***ADÜSEM EU Erasmus + Writing a Project Education- 10.06.2017*****SAP Material Management Intensive Introduction Course****Sagesse Consultancy- December 2013****CEVİZ INFORMATION TECHNOLOGIES -Sharepoint Course - January 2014****ERASMUS + TEACHING STAFF MOBILITY**

10.04.2017 – 15.04.2017 (8 hours of Human-Computer Interaction lessons have been instructed.)

SMK UNIVERSITY OF APPLIED SOCIAL SCIENCES- Lithuania - Vilnius**PROJECTS AND EXTERNAL ACTIVITIES:**

T.C SÖKE DISTRICT GOVERNORSHIP “CHESS FOR EVERYONE EVERYWHERE” PROJECT – Adnan Menderes University Project Coordinator
(September 2017 -)

Adnan Menderes University Söke Faculty of Business **“YOUNG IDEAS”** Student Club
Advisor
(September 2017 – June 2018)

Adnan Menderes University Söke Faculty of Business **“CULTURAL AMBASSADOR”**
(April 2017 -)

Adnan Menderes University Söke Faculty of Business **“FACULTY WEBSITE RESPONSIBLE”**
(January 2016 -)

