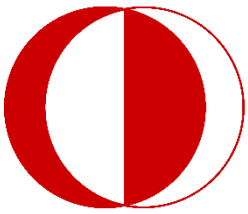
Middle East Technical University

Department of Statistics



STAT365

SAMPLING AND SURVEY TECHNIQUES

TERM PROJECT

Language Learning Process of METU Students

Aybüke Altuntaş	2361095
Gözde Akkaya	2361061
Gülin Oğuz	2429181

January 2022

TABLE OF CONTENT

1. INTRODUCTION

Learning a foreign language provides numerous advantages, including increased job chances, improved cognitive capabilities, higher test scores, and a deeper understanding of cultures and nations. When people who speak different languages need to communicate, they have a number of options, the most obvious of which is to learn and teach a second language. Learning a language is a recursive and ongoing process as we embrace and navigate new learning and contexts. ("English Language Arts Curriculum Framework: A Living Document", 2020) and just like any other type of learning, language learning process of an individual can be influenced by many factors. In this study, factors such as gender, reasons, beginning time, amount of time spared for language learning and more will be examined in order to make conclusions about the main and minor objectives.

"Language is the expression of ideas by means of speech-sounds combined into words. Words are combined into sentences, this combination answering to that of ideas into thoughts."

The American linguists Bernard Bloch and George L.

2. AIM OF RESEARCH

2.1. Main Objective

The target population mainly consists of METU students who have started learning a foreign language in the early stages of their lives, however most of them still face trouble whenever they perform an action that requires them to make use of their foreign language skills. This situation yielded the question "How does language learning of METU students process?". Thus, the decision to investigate the process of language learning and the effects of different factors was made.

2.2. Minor Objectives

- Investigating the relationship between the participants' faculty and their average time of studying a language
- Finding out the relationship between the gender and the beginning time of their language learning adventure.
- Examining whether listening to the music is beneficial for language learning or not. If it is beneficial, findings out the aspects it contributes to while learning.
- Investigating which skill is the most focused for students to learn/improve for each given foreign language in our survey.
- Finding out the main reason(s) why the foreign languages given in our survey is learned.

2.3 Research Philosophy

The realist approach was opted for this research because realist evaluation (data collection, interpretation and analysis) and realist synthesis (synthesis of evidence or literature) were going to be made use of. Additionaly, the target population was defined as METU students so a context for the study would be present and the survey questions could be prepared to measure interactions and relative efficiencies of both material and languages for each individual. Moreover, realistic approach is context sensitive so the conclusion of our research is not generalizable, it is open to change.

3. SURVEY METHODOLOGY

3.1 Survey Design

3.1.1. Data Collection

In order the collect data, an online questionnaire consisting of 19 questions was prepared on Google Forms and distributed to the target population, METU students, via online platforms such as messaging services and social media. The questionnaire included open-ended, multiple choice and Likert scaled questions. After the two week period of gathering replies for the survey, the data was extracted from Google Forms and converted to an Excel table and a CSV file. Microsoft Excel and R was used in the analysis and visualization of the data.

3.1.2. Sample Design

Originally, there were 234 replies in total for the questionnaire. After the removal of irrelevant replies (e.g. mocking attitude, faculties that are not in METU) the number of replies dropped to 228. Since the questionnaire was distributed via online platforms among METU students, it was forwarded from one student to another. As a result of this, the sampling of the research is snowball sampling.

3.2 Methods of Analysis

To analyze the data further, descriptive statistics and some statistical tests are used by using R Programming (R Studio) and Microsoft Office Excel.

3.2.1 Descriptive Statistics

For visualization and identification analysis of data, the following graphs were obtained: Pie charts, Bar charts (Frequency Distributions), Tables, Statistical Plots (Q-Q Plot), and Stacked Bar Plot.

3.2.2 Statistical Tests

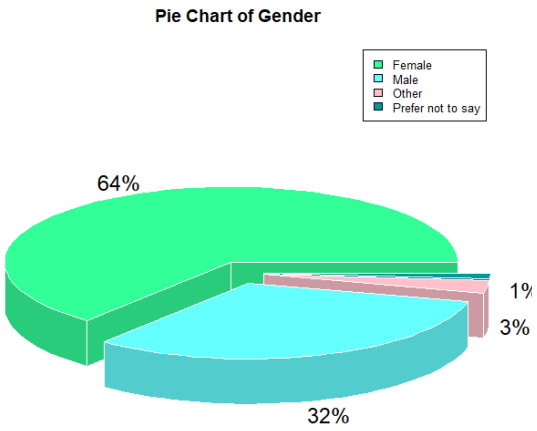
Pearson's chi-squared test is considered since it is a statistical test applied to sets of categorical data to determine how probable any apparent difference between the sets is due to chance. Also, in the purpose of using ANOVA Test, Shapiro-Wilk Normality Test and Variance-Equality Test were considered.

4. DATA DESCRIPTION

The data contains 228 observations and 37 variables. The data is in the data frame format. The variables are Birth Year, Gender, Entrance Year, Department, Faculty, Interaction, Beginning Time, Reason, Most Focused Skills, Difficulty in Reading, Difficulty in Listening, Difficulty in Speaking, Difficulty in Alphabet, Difficulty in Grammar, Difficulty in Vocabulary, Study Hours per Week, Materials, Courses School, Face to Face Language Courses, Online Courses, Tutor, Mobile Apps, Social Media, Streaming Services, Textbooks, Other, Video Watch Language, English Level Reading, English Level Listening, English Level Writing, English Level Speaking, Resources Language, Listening Music, Music Efficient, Music Skills.

Gender

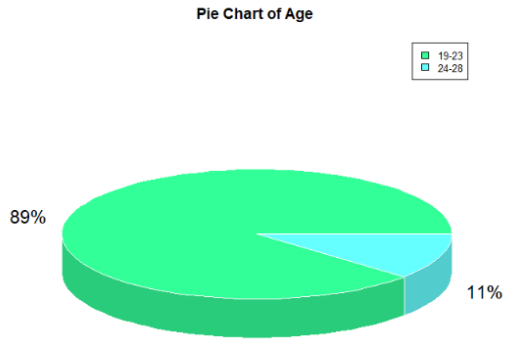
Figure 1: Pie Chart of Gender



The pie chart on the left represents the gender distribution of respondents. It can clearly be interpreted that female students make up the majority of the data and with a rate of 32%, males are the second largest group. 146 people chose the female option, 73 people chose the male option, 6 people chose the other option and 3 people chose the prefer not to say option

Age

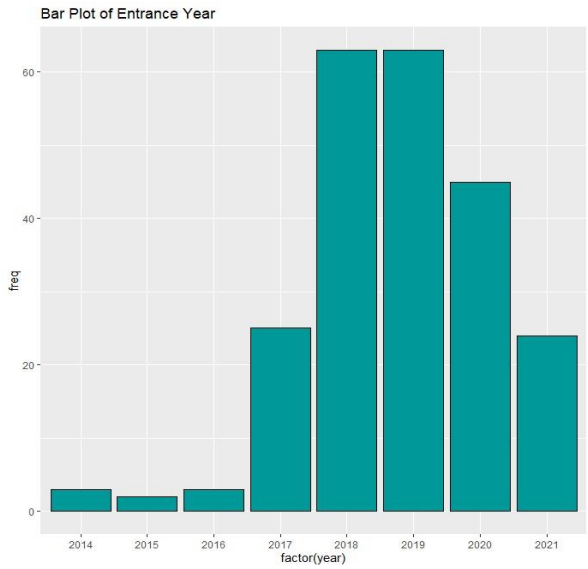
Figure 2: Pie Chart of Age



From the pie chart of age, it is apparent that 89% of students' ages are between 19 and 23. The rest of the students' (11%) ages are between 24 and 28.

Entrance Year

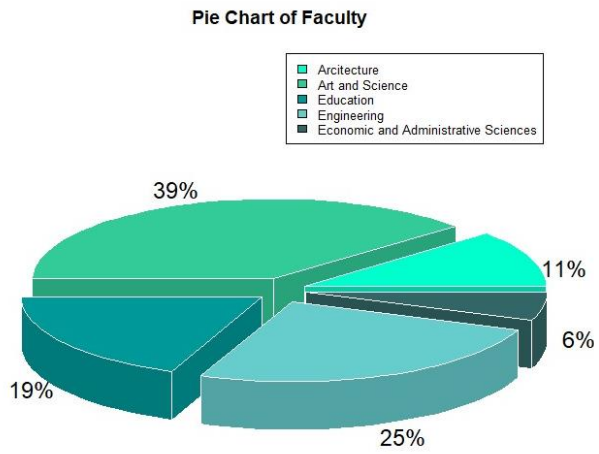
Figure 3: Bar Chart of Age



The bar chart on the left represents the entrance year of the respondents. Most of the students who participated in this study have entered METU in 2018 and 2019 (higher than 60 respondents). The number of students who have entered METU in 2020 also have a significant participation rate when compared to previous years, such as 2016. The students who entered in 2014, 2015, and 2016 are also respondents of this study, but they do not make up the majority of the research.

Faculty

Figure 4: Pie Chart of Faculty



The pie chart of faculty displays METU students' faculties who are respondents to this survey. It can be seen from the pie chart that the majority of the students (39%) belong to the Faculty of Arts and Science. In this study, 88 of the students belong to the Faculty of Arts and Science, 58 of them belong to the Faculty of Engineering, 43 of them belong to the Faculty of Education, 26 of the students' belong to the Faculty of Architecture, and 13 of the students' belong to the Faculty of Economics and Administrative Sciences.

5. DATA ANALYSIS, FINDINGS AND FURTHER DISCUSSIONS

5.1. Finding out the main reason(s) why METU students are learning the foreign languages.

Hypothesis 1: Is there a characteristic association between the languages and their reasons for learning?

In order to investigate the learning reasons, a multiple-choice question that enabled the respondents to choose more than one option was asked on the questionnaire. (e.g., a student could choose both for educational purposes and for travel).

	Learning Reasons						
Languages that are learned		For Work	For Educational Purposes	For Immigration	For Travel	Out of interest, as a hobby	For a foreign person in the family or close circle
	English	141	182	94	110	125	9
	German	42	47	36	32	47	3
	Spanish	18	29	16	18	29	3
	French	11	19	10	8	17	1
	Korean	9	15	13	14	20	1

Table 1: *The relationship between the most preferred languages and their learning reasons*

According to Table 1, the percentage of most frequent three choices of reasons:

English	• "For work" choice is 21. • "For educational purposes" choice is 28. • "Out of interest, as a hobby" choice is 19.
German	• "For work" choice is 20. • "For educational purposes" choice is 22. • "Out of interest, as a hobby" choice is 22.
Spanish	• "For work" choice is 16. • "For educational purposes" choice is 26. • "For travel" choice is 16. • "Out of interest, as a hobby" choice is 26.
French	• "For work" choice is 17. • "For educational purposes" choice is 28. • "Out of interest, as a hobby" choice is 26.
Korean Language	• "For educational purposes" choice is 20. • "For travel" choice is 19. • "Out of interest, as a hobby" choice is 27.

The relationship between these two variables was examined by Pearson Chi-squared Test.

H_0: The two variables are independent.

H_1: The two variables are related to each other.

According to the test; p-value was found as 2.2e-16 which is lower than the significance level (alpha=0.05). Thus, the null hypothesis can be rejected. In other words, there is significant evidence to state that there is an association between the languages that are learned and their reasons.

5.2. Investigating which skill is the most focused for students to learn/improve for each given foreign language in our survey.

Language-learning Skills (LLs)								
Most preferred Languages		Listening	Writing	Speaking	Vocabulary	Grammar	Reading	Alphabet
	English	138	81	115	95	65	127	21
	German	46	21	39	34	31	32	5
	Spanish	24	11	23	22	11	23	4
	French	15	6	14	13	8	11	4
	Korean	16	6	11	15	8	13	6

Table2: The amount of selection of each LLs for five most preferred languages among METU students.

According to our data, the most preferred five languages that METU students are learning are English, German, Spanish, French, and Korean. In the purpose of getting which language-learning skills are most preferred for each language, it is investigated in table 2. According to table 2, Students who are learning English mainly focus on listening (%22), speaking (%18), and reading (%20). Students who are learning German mostly focus on listening (%22), speaking (%19), and vocabulary (%16). Students who are learning Spanish mostly focus on listening (%20), speaking (%20), and reading (%19). Students who are learning French mostly focus on listening (%21), speaking (%20), and vocabulary (%21). Students who are learning Korean mostly focus on listening (%21), reading (%18), vocabulary (%20). As a result, it can be concluded that listening, reading, speaking, and vocabulary are the most focused skills that METU students want to advance since they highly concentrate on these skills while studying a language they are learning.

5.3 Is there a relationship between average hour of studying per week and the faculties of METU students ?

To examine the relationship, the initial examination is to check the normality of the variables. For this purpose, Shapiro-Wilk Normality Test is used with $\alpha = 0.05$ on “faculty” variable.

H_0: The data are normally distributed.

H_1: The data are not normally distributed.

The results obtained after performing the test are as follows: The data are normally distributed. It can be also investigated from Q-Q Plot in the below.

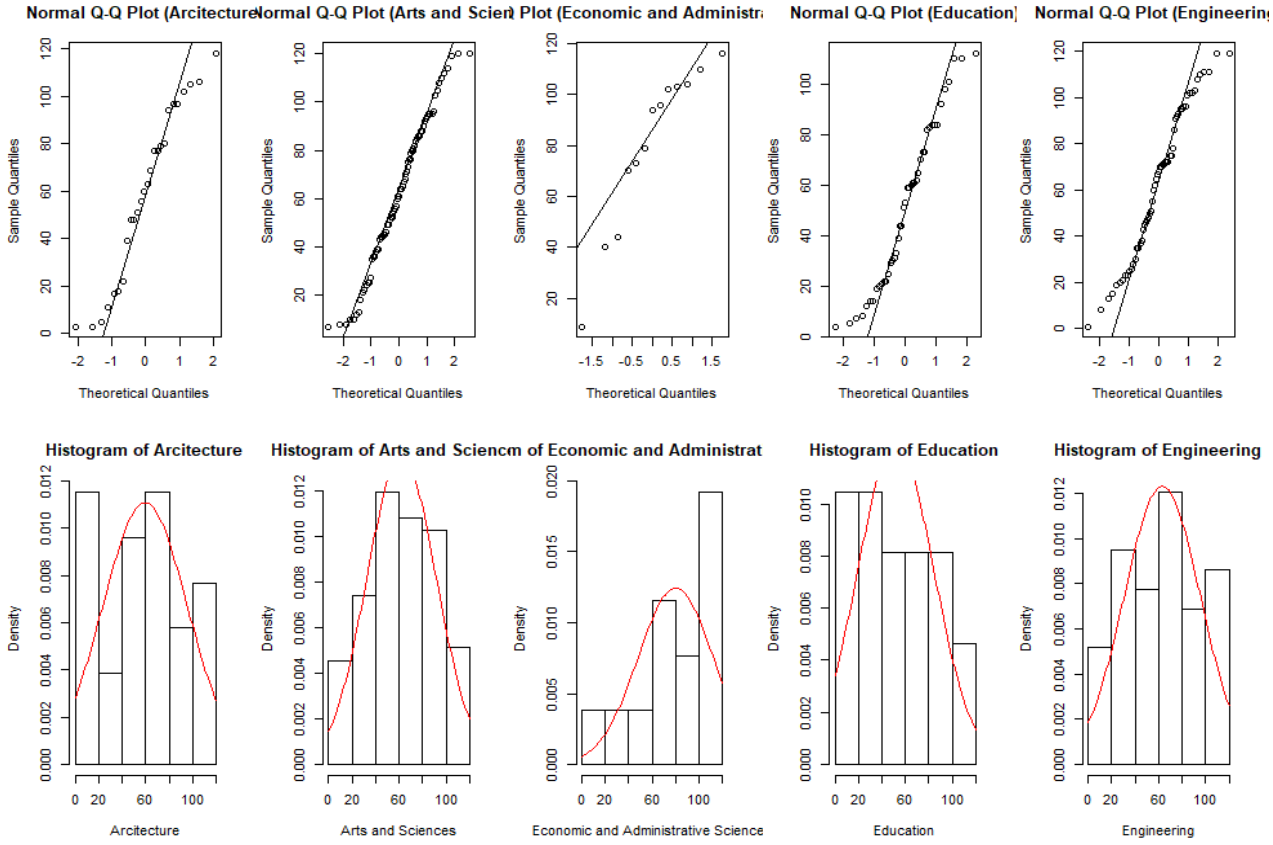


Figure 5: Normal Q-Q Plot and Histogram of Faculty based on Average Study Hour Per Week

To apply ANOVA Test, variance-equality check is considered. After the variance equality test, it is shown that variances are homogeneous. Since ANOVA Test assumptions are satisfied, ANOVA test is applied.

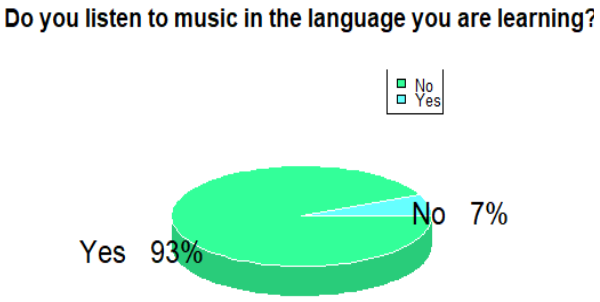
H_0: There is no difference in means.

H_1: There is a difference in means.

According to ANOVA Test results, since the p-value of residuals (0.0951) is larger than the threshold value (0.05), the null hypothesis cannot be rejected. This shows that the means are not statistically different than each other. Therefore, the faculties of METU students have no direct effect on average hour of studying per week.

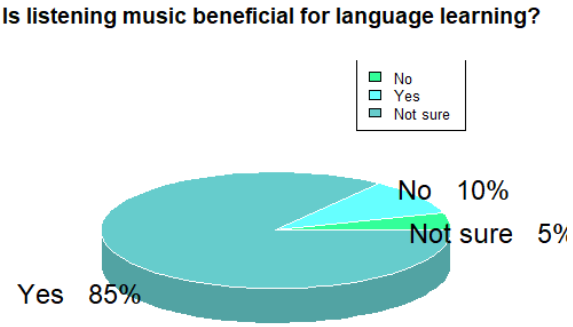
5.4. Examining whether listening to the music is beneficial for language learning or not. If it is beneficial, finding out the aspects it contributes to while learning.

Figure 6: Pie Chart of listening to music



According to the pie chart which is on the left side, most of the participants (%93) listen to music in the language they are learning.

Figure 7: Pie Chart of opinions on the benefits of music for LL



Most of the participants who listen to music in the language they are learning think that listening to music in the language they are learning can be beneficial for language learning.

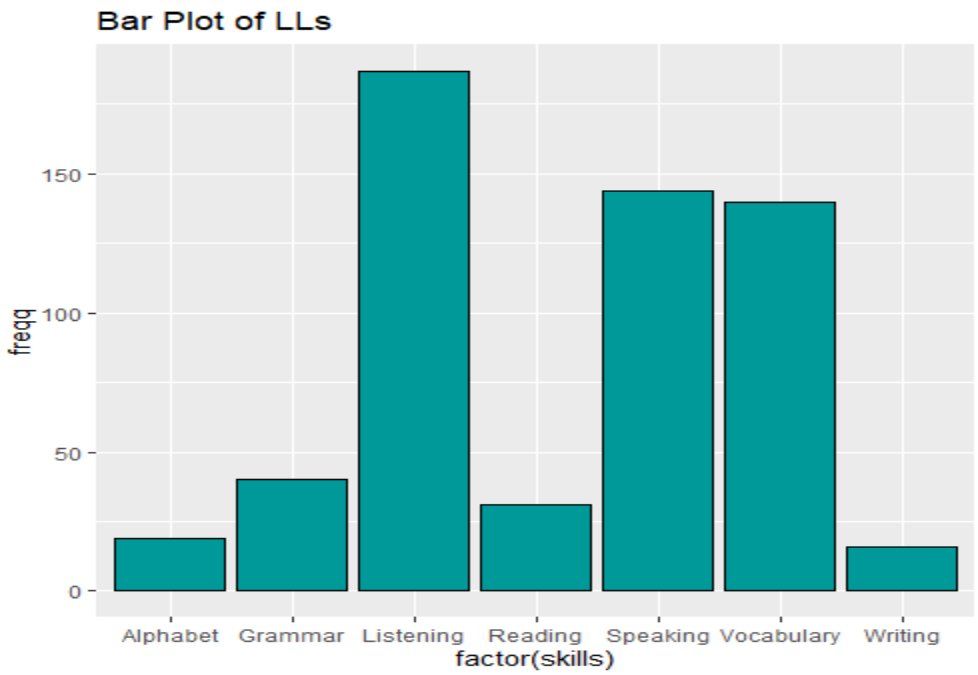


Figure 8: Bar Plot of LLs vs Frequency

In this bar plot of LLs versus their frequency of selection can be investigated. According to the bar plot, most of the METU students who believe that listening to music is beneficial for language learning think that listening to music in the language they are learning can especially contribute to listening skills, speaking skills, and vocabulary knowledge. It is also a substantial selection for other language skills such as reading and writing.

5.5. Finding out the relationship between the gender and the beginning time of their language learning adventure.

The main aim is to investigate whether there is an association between beginning time of their language learning journey, and gender. To examine this relationship further, Pearson Chi-squared test were performed.

H_0: The two variables are independent.

H_1: The two variables relate to each other.

The p-value which was equal to 0.152 is clearly higher than the significance level $\alpha = 0.05$. Therefore, the null hypothesis cannot be rejected. This means that there is no significant evidence to say that gender of students and participants' beginning time of language learning are dependent.

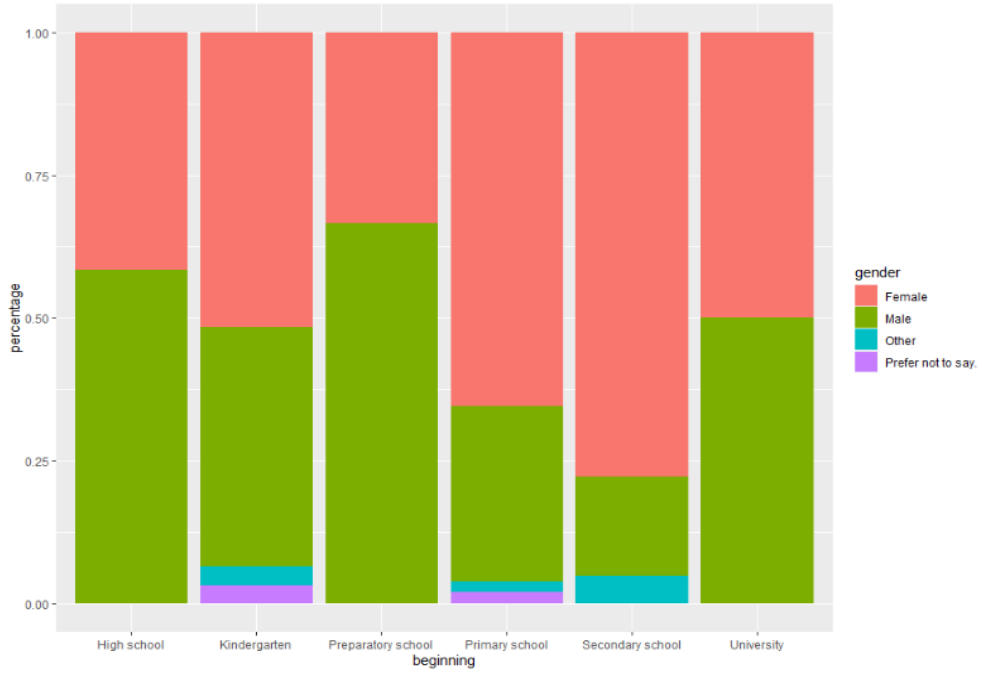


Figure 9: Stacked Bar Plot for the percentages of Beginning vs. Gender

To clarify, Figure 9 horizontal axis represents the beginning time, and the vertical axis means percentages of gender variables. As an evidence in the above plot, there is no association between gender and beginning time. Thus, The plot supports the Pearson Chi-squared test result.

6. CONCLUSION AND FURTHER DISCUSSIONS

Learning a language is a time-consuming, complex endeavor that demands dedication, perseverance, and hard effort. As a result, it would be inappropriate to look at the language learning process from simply one angle. In this study, language learning habits of METU students are considered for this aim. Questions such as if they listen to music in the language they are learning or not, when they started their language learning adventure, which LLs skills they want to develop and focus on, how many hours per week they spent studying, and so on were asked in order to analyze their behaviors. In addition, categorical characteristics such as gender and department are retrieved. We asked, despite the fact that we did not evaluate certain other variables. Our initial findings in that regard were not surprising. According to Riehl (2016), a person who begins learning a second language at a young age will be able to acquire a foreign language more effectively in the future. He also stated that acquiring a language generates a system in the minds of multilingual youngsters that encompasses all languages. METU students who know English are good in other languages because they know how to study and what to study, according to our agreement. Horn's work reveals that music is a sort of discourse that employs tones, pitch, and rhythm as a universal language. Music also expands one's vocabulary and teaches articulation and speech, according to her. She emphasizes that music plays an important part in learning a second language, highlighting that the learner must have well-developed listening abilities in order to "be able to study, interpret, and experience music and language." We wanted to understand more about how students felt about the connection between music and language learning. Our findings are similar to Horn's, as they believe that listening to music in the language they are learning is a very effective way to learn the language. The four "macro skills" (listening, speaking, reading, and writing), according to Xue Ping (1997) in The Internet TESL Journal, are all required for language acquisition and use. Any language cannot be learned in isolation; we must blend the four LLSs to become proficient users (language-learning skills). To have a better understanding of this study, it will be investigated which LLs are the most concentrated among METU students and which LLs are the most difficult to learn. Our findings revealed that when learning a language, students mostly concentrate on these four macro abilities, particularly listening and speaking. They, on the other hand, have difficulties speaking and listening.

7. REFERENCES

English language arts curriculum framework (2020). The Government of Manitoba. Retrieved from

https://www.edu.gov.mb.ca/k12/cur/ela/framework/full_doc.pdf

Horn, C.A. 2007. English second language learners: using music to enhance the listening abilities of grade ones. Masters thesis. Unisa.

http://etd.unisa.ac.za/ETDdb/theses/available/etd09212007.133117/unrestricted/dissertation.pdf.

Riehl, Amanda B., "Americans' Foreign Language Deficit and Possible Solutions" (2016). *HON499 projects*. 8.

https://digitalcommons.lasalle.edu/honors_projects/8

Trager, George L., Bloch, Bernard. (1942). *Outline of Linguistics Analysis* . Baltimore: Linguisic Society of America.

Xue-Ping, G. (1997). A scheme for the obtaining of language skills. The Internet TESL Journal, 3(6).

8. APPENDIX

Language Learning Process of METU Students

Questionnaire

1) **Please enter your year of birth.**
2) **Please chose your gender**.
   i) Female
   ii) Male
   iii) Other
   iv) Prefer not to say
3) **Please indicate your university entrance year.**
4) **Please indicate your department; if you are a graduate student, please indicate the department you graduated from. (STAT, CHEM, ...)**
5) **What language(s) are you currently learning?**
   i) English
   ii) German
   iii) French
   iv) Spanish
   v) Italian
   vi) Japanese
   vii) Korean language
   viii) Mandarin (Chinese)
   ix) Russian
   x) Greek
   xi) Other
6) **Have you interacted previously with the language(s) you are learning?**
   i) In the early stage of education (e.g., high school, secondary school, ...)
   ii) Being abroad (e.g., language school, residing abroad, ...)
   iii) Out of interest, as a hobby.
   iv) Cultural activities (e.g., music, cinema, literature, ...)
   v) I did not have any interaction.
   vi) Other
7) **When did your adventure of learning a foreign language begin?**
   i) Kindergarten
   ii) Primary school
   iii) Secondary School
   iv) High School
   v) University
8) **Why are you learning this language(s)?**
   i) For work
   ii) For immigration
   iii) For travel
   iv) Out of interest, as a hobby
   v) For educational purposes
   vi) For a foreign person in the family or close circle
   vii) Other
9) **Which areas do you focus on most while studying the language you learn?**
   i) Writing
   ii) Reading
   iii) Listening
   iv) Speaking
   v) Alphabet
   vi) Grammar
   vii) Vocabulary
   viii) Other
10) **How much difficulty do you have while studying the areas you choose in the previous question? Please rate it. (Rate the above areas. (Lickert scale))**
   i) Extremely
   ii) Moderately
   iii) Not at all
11) **On average, how many hours per week do you spend learning the current language(s)?**
   i) I do not devote time to learning.
   ii) 1-30 minutes
   iii) 31-60 minutes
   iv) 61-90 minutes
   v) 91-120 minutes
   vi) >120 minutes
12) **What learning methods/materials do you use when learning a language? Please select all that apply.**
   i) Courses given by school
   ii) Face to face language course
   iii) Online language course (e.g. Udemy, Coursera and etc.)
   iv) Tutor
   v) Mobile apps (e.g. Duolingo, Rosetta Stone and etc.)
   vi) Social media (e.g. YouTube, Instagram, and etc.)
   vii) Streaming services (e.g Netflix, Amazon Prime and etc.)
   viii) Textbook(s)
   ix) Other
13) **Rate the learning methods/materials you use in terms of efficiency. (Rate the above methods/materials. (Lickert scale))**
   i) Good
   ii) Fair
   iii) Poor
14) **How do you usually prefer to watch a video/movie/series in the language you are learning?**
   i) Dubbing in your native language
   ii) English subtitle
   iii) Subtitled in the language you are learning
   iv) No subtitles

v) Or in another language (except English.) (If you selected this option, specify the language in the 'Other' section.)
vi) Other

(If you don't know English, skip the 15th and 16th questions.)

15) Rate your English skills. (Lickert scale with Good-Fair-Poor options.)
i) Reading
ii) Listening
iii) Writing
iv) Speaking

16) In which language are the resources you use when learning a language other than English?
i) In the language I learned
ii) English
iii) In my native language
iv) I am not learning any language other than English.

17) Do you listen to music in the language(s) you are learning?
i) Yes
ii) No

18) Do you think listening to music is beneficial for language learning?
i) Yes
ii) No

(If you answered "Yes" to the previous question, answer the following question, otherwise you can finish the survey.)

19) Which skills do you think listening to music contributes to?
i) Writing
ii) Reading
iii) Speaking
iv) Listening
v) Grammar
vi) Alphabet
vii) Vocabulary
viii) Other

```
inst_pack_func <- function(list.of.packages){new.packages <-
list.of.packages[!(list.of.packages %in%
installed.packages()[,"Package"])]if(length(new.packages))
install.packages(new.packages)lapply(list.of.packages,function(x){library(x,character.only=T
RUE)})}list.of.packages <-
c("MASS","tidyverse","dplyr","ggplot2","knitr","DAAG","BSDA","onewaytests","readxl","p
lotrix","stringr");inst_pack_func(list.of.packages);data365 <- read_excel("STAT_365
DATA.xlsx")
;summary(data365);colnames(data365);view(data365);table(data365$gender);colors_gender
<- c("#33FF99","#66FFFF","pink","#009999");gender_labels <-
c("Female","Male","Other","Prefer not to say");x_gender <-c(146,73,6,3);piepercent_gender
<- round(100*x_gender/sum(x_gender));final_labels_gender <-
paste(piepercent_gender,'%',sep =
"");final_labels_gender;pie3D(x_gender,labels=final_labels_gender,explode=0.1,main="Pie
Chart of Gender ",col=c("#33FF99","#66FFFF","pink","#009999"), labelcol = "black", border
= "white");legend("topright", c("Female","Male","Other","Prefer not to say"), cex = 0.8,fill
=colors_gender);table(data365$ent_year);year <- c("2014", "2015", "2016", "2017", "2018",
"2019", "2020", "2021");freq <- c(3,2,3,25,63,63,45,24);data_year <-
as.data.frame(year,freq);ggplot(data_year, aes(x=factor(year),
y=freq))+geom_bar(stat="identity",fill="#009999",colour="black")+ggtitle("Bar Plot of
Entrance Year");table(data365$faculty);colors_faculty <-
c("#00FFCC","#33CC99","#009999","#66CCCC","#336666");faculty_labels <-
c("Arcitecture", "Art and Science","Education", "Engineering","Economic and Administrative
Sciences");x_faculty <- c(26,88,43,58,13);piepercent_faculty <-
round(100*x_faculty/sum(x_faculty));final_labels_faculty <- paste(piepercent_faculty, "%",
sep =
"");final_labels_faculty;pie3D(x_faculty,labels=final_labels_faculty,explode=0.1,main="Pie
Chart of Faculty ",col=c("#00FFCC","#33CC99","#009999","#66CCCC","#336666"),
labelcol = "black", border = "white");legend("topright", c("Arcitecture", "Art and
Science","Education", "Engineering","Economic and Administrative Sciences"), cex = 0.8,fill
=colors_faculty);colors_age <- c("#33FF99","#66FFFF");x_age <-  c(202,26);labels_age <-
c("19-23","24-28");piepercent_age <- round(100*x_age/sum(x_age));final_labels_age<-
paste(piepercent_age,'%',sep = "");pie3D(x_age,labels=final_labels_age,  main="Pie Chart of
Age ",col=c("#33FF99","#66FFFF"), labelcol = "black", border = "white");legend("topright",
c("19-23","24-28"), cex = 0.8,  fill =colors_age); data365$mınutes <-sample(c(1:120),size =
228 , replace =
TRUE);data365$faculty;library("onewaytests");nor.test(mınutes~faculty,data=data365);krusk
al.test(data365$mınutes,data365$faculty);homog.test(mınutes~faculty, data =
data365);aov(mınutes~faculty, data = data365);summary(aov(mınutes~faculty, data =
data365));chisq.test(data365$language,data365$reason);chisq.test(data365$gender,data365$b
eginning);ggplot(data365, aes(x = beginning)) + geom_bar(aes(fill = gender), position = 'fill')
+
labs(y="percentage";table(data365$gender,data365$beginning);table(data$listening_music);m
usic <- c("No","Yes");freq_s <- c(16,212);percentage_skill<-
round(freq_s/sum(freq_s)*100);labels_music<paste(music,",percentage_skill);labels_music;fi
nal_labels_music<-
paste(labels_music,'%',sep="");final_labels_musicpie3D(freq_s,labels=final_labels_music,
```

```
main="Do you listen to music in the language you are
learning?",col=c("#66FFFF","#33FF99"), labelcol = "black", border = "white");colors_music
<- c("#33FF99","#66FFFF");legend("topright", c("No","Yes"), cex = 0.8, fill
=colors_music);table(data$music_efficient);efficient <- c("Not sure","No","Yes");freq_eff <-
c(11,23,194);percentage_efficient <-round(freq_eff/sum(freq_eff)*100);labels_efficient <-
paste(efficient,",percentage_efficient);labels_efficient;final_labels_efficient <-
paste(labels_efficient,'%',sep="");final_labels_efficient;pie3D(freq_eff,labels=final_labels_eff
icient,main="Is listening music beneficial for language learning?
",col=c("#33FF99","#66FFFF","#66CCCC"), labelcol = "black", border =
"white");colors_efficient <- c("#33FF99","#66FFFF","#66CCCC");legend("topright",
c("No","Yes","Not sure"), cex = 0.8,fill = colors_efficient);data_music <-
read_excel("musicskills.xlsx");data_music;table(data_music$music_skills);skills <-
c("Listening", "Speaking", "Writing", "Reading", "Vocabulary", "Grammar",
"Alphabet");freqq <- c(187,144,16,31,140,40,19);data_musicc <-
as.data.frame(skills,freqq);data_musicc;ggplot(data_musicc, aes(x=factor(skills),
y=freqq))+geom_bar(stat="identity",fill="#009999",colour="black")+ggtitle("Bar Plot of
LLs")
```