

Background Music Will not Significantly Distract Attention*

Hong Pan

April 25 2021

Abstract

Music has gradually become an essential part of our lives, and it has become a common trend for most students to choose to listen to music while studying. This paper used the dataset collected by Arriken Worsley(Worsley 2020) to examine whether the background music affects students' attention level by conducting a within-subjects experiment to compare the time of finishing the Trail Making Test with and without music. The linear mixed model results show that background music does not significantly impact attention level but has a significant negative relationship with the order of the test being done.

Keywords: Background Music, Attention, Trail Making Test, Executive Functioning Skills

Contents

1	Introduction	2
2	Data	3
2.1	Participants	3
2.2	Experiment	3
2.3	Mearues	4
2.4	Descriptvie Analysis	5
3	Model	8
3.1	Linear Mixed Model	9
3.2	Modeling Procedure	9
3.3	Model Diagnostics	10
4	Results	12
4.1	Random Effect Result	12
4.2	Full Model Result	12
4.3	Model Assessment Result	14

*Code and data are available at: https://github.com/Data815/background_music

5	Discussion	14
5.1	Ethics statement	14
5.2	Descriptive Analysis Results	14
5.3	Model Results	15
5.4	Practical Implications	15
5.5	Weaknesses and Next Steps	16
6	Appendix	18
6.1	Test 1	18
6.2	Test 2	19
6.3	Survey	20
6.4	Final Model Output	20
	References	22

1 Introduction

Should students listen to music while studying? In contemporary times, music has become an indispensable part of human life. With the development of technology, students can more easily and conveniently access music through various devices and channels. According to Nielsen’s report, teens use an average of 3.8 devices in a week to engage with music, such as smartphones, laptops, and tablets, and on average, teens spend over 32 hours per week listening to music(Nielsen 2017). Meanwhile, the survey took by Jackson Gutknecht illustrates that around 75 of 80 students said they listen to music while studying(Gutknecht 2018). Listening to music during the learning process has become a common trend for many students, therefore, it is important to understand the influence of background music on students’ attention level.

Trail Making Test, which is used to measure executive functioning skills, was utilized to study whether background music would influence students’ attention level. Executive functions refer to the mental processes needed when people have to concentrate or pay attention, which is “a specific set of attention-regulation skills involved in conscious goal-directed problem-solving.(Philip David Zelazo 2016)” Defectiveness of executive functioning skills will make students have a hard time paying attention(Team 2021). In the Trail Making Test, each round consists of several dots numbered from 1 to 12 and lettered A to J, randomly distributed on a page. The participants are required to connect the circles as soon as possible.¹ The paper written by Arielle S. Dolegui(Dolegui 2013) used repeated measures to design experiments to study the impact of different types of music and playing volume levels on students’ cognitive ability to complete assignments. The result illustrates that students perform the best in silence. Meanwhile, the paper published by Janina and Tina investigates(Janina A. M. Lehmann 2017) how background music influences learning by conducting a between-subjects design experiment. The result for their paper demonstrates that background music did no influence performance for learners. Is it true that background music does not affect the ability of students to process information?

The within-subjects experiment is conducted by randomly selected 47 students from Elizabeth Blackburn School of Science, and each participant needs to complete two rounds of tests, with and without music. In this study, the practice effect² may exist since the students need to do the test twice continuously. In order to reduce the impact from practice effect, participants were randomly selected to two different groups with different starting order. One group started with a test with background music, while the other group started

¹The order should alternate between numbers and letters, i.e. 1, A, 2, B,...

²People improve their scores simply by taking the same test multiple times. (Kendra Cherry,2019)

the test in silence. The dataset created by Arriken Worsley(Worsley 2020) records the time of accomplishing the test, gender of the student, music favourability, and the order of test which is starting with the music or without the music. The descriptive analysis results show that students performed better without music. In addition, whatever the order of the test, the second round will be faster on average. Furthermore, background music has the biggest influence on male’s executive functioning skills. Because the dataset is longitudinal, the linear mixed model is applied to do the statistical analysis. The result from the model illustrates that the students’ attention level is not significantly impacted by listening to the music. The paper concludes that the background music does not significantly impact students’ attention level.

The rest of this paper is structured as follows. In the data section, the experimental process, sampling method and data set variables will be explained. The descriptive analysis detail will also be presented. In the model section, the reason for choosing the model, model assumptions, and model selection procedure will be discussed. In the result section, the statistical analysis detail and model validation result will be illustrated and explained. In the discussion section, the main findings, limitations, ethics statements, and further research will be discussed. In the Appendix, the screenshot for the survey and the real Trail Making Test used for the experiment will be shown. Finally, the interactive graphs display on Shinyapp³ allow the audience to explore the data distribution by gender, test type and rating.

2 Data

The data used in this paper analyzed through the R statistical programming language(R Core Team 2020) to investigate whether background music affects students’ level of attention. The Tidyverse package(Wickham et al. 2019) for data processing, the KableExtra package(Zhu 2021) and plyr package(Wickham 2011) for generating tables, and the ggplot2 package(Wickham 2016) for drawing graphs are used during the analysis process. Meanwhile, the here package(Müller 2020) for importing datasets and the lme4 package(Bates et al. 2015) for creating models are also used. The performance package(Lüdecke et al. 2020), lattice package(Sarkar 2008) and MuMIn package(Bartoń 2020) are used to draw charts, check model assumptions and model validity. Finally, the stargazer package(Hlavac 2018) is applied to visualize the model result.

2.1 Participants

The researcher experimented in 2020 at the Elizabeth Blackburn School of Science, a secondary school funded by the Melbourne government. The study sample comprised 47 students between 13 and 17 years old randomly chosen from the school. Both male and female students participated in the research on whether background music affects people’s attention. Figure 1, the pie chart, illustrates that 51% are female, 47% are male, and the rest 2% are transexual.

2.2 Experiment

2.2.1 Intervention

The purpose of the experiment is to examine whether background music affects the attention of students. The design of the experiment is to randomly select some students from the school to complete the “Trail Making Test” with and without background music and record the time of completion. For this experiment, the intervention is the presence of background music. Experiment designer Arriken Worsley wants to determine whether background music will affect students’ attention level by studying whether students take longer to complete the test with background music than without background music. If the result is that the time for the student to complete the test with background music is longer than the time without background music, it means that the background music affects the student’s attention to a certain extent. Otherwise, background

³The Shinyapp link for the interactive graphs is https://hongtrista.shinyapps.io/Shinyapp/?_ga=2.86381278.759236858.1619203106-1272735195.1619203106

music will not affect students' attention. The experiment can help the target audience develop some basic understanding of whether music impacts attention and correctly decide whether to listen to music in future studies or avoid inefficiency.

2.2.2 Procedure

As the paper aims to understand whether background music will affect students' attention level, all students in the world would be considered as the target population. The frame population is all students from the school. 47 students being randomly selected from the school as a sampled population to finish the experiment. The detailed procedure will be explained as followed.

The 47 students chosen randomly from the school were randomly divided into two groups to start the Trail Making Test in a different order. 27 students started the experiment in a quiet environment, and the remaining 20 students started with the existence of background music. Figure 8 and Figure 9 in the appendix show two different Trail Making Test used in different test environments during the experiment. Each test contained 23 randomly distributed dots, including 12 numeric points from 1 to 12 and 11 letter points from A to K. Participants must start connecting from 1 and connected these points to the final point in alternating ascending order, such as 1, A, 2, B, 3, C, et cetera. During the process, the line should not cross over. The background music, the intervention, played during the experiment would be Designer's song, Panda, because, during the experiment period, it was at the top of the billboard and was the most popular music.

The experiment was conducted in the form of a class. Students selected to the same group started the test uniformly, and a timer was placed in front of each student to record the completion time. In order to counteract the learning effect gained from the first test to the second one to some extent, changing the order of starting the experiment was applied. The students in the first group started the Trail Making Test without background music, then did another test while playing the background music, Panda. The second group of students started to do the test with background music and did the second test in silence for control. After each round, the accomplishing time was recorded in unit second. Meanwhile, when participants finished all the two tests, they were required to fill in an anonymous survey(Figure 10) shown in the appendix that asked for feedback, gender, and personal enjoyment of the song played, on a scale from one to five.

2.3 Mearues

The paper intends to use the dataset collected from the experiment set up by Arriken Worsley(Worsley 2020) to analyze whether the background music will influence students' attention level. The dataset records 94 observations with 13 variables. The data set contains some empty columns and variables that are not related to the research topic⁴; therefore, only 8 variables and 94 observations are left after changing the dataset into a longitudinal.

The **Id** represents a student, and a number represents a student. **Test_Type** illustrates whether the students finished the test with background music or without, which indicates the data comes from the control group or the treatment group. **Time_Sec** shows how many seconds does the student to accomplish the Trail Making Test. The column, **Gender**, shows the gender of the student. **Rating** indicates the student's preference for background music. After completing the experiment, students need to fill out a questionnaire(Figure 10 in appendix) anonymously, including rating background music from 1 to 5. The higher the score, the more they like the background music. The **Feedback** column shows the students' feelings and evaluations of the entire experiment. **Change** shows test 2 compares with test 1, by what percentage increase or decrease the time. **Order** records the order of the test, which helps to illustrate whether the test is the first or second round.

⁴The variables are "question" (the experimenter's comment) as well as "Diff-AVG" and "DIF_AVG". The raw datasets before cleaning could be found in the Input folder, music_first.csv and no_music_first.csv.

2.4 Descriptvie Analysis

2.4.1 Gender distribution

According to the experiment, students are randomly selected from the school. Students need to fill out a questionnaire after completing two “Trail Making Test” to inform the experiment initiator of their gender and other information. The pie chart(Figure 1) shows that among the students selected for the experiment, the proportions of men and women are almost the same, which leads to the conclusion that the distribution of men and women is almost the same. Female accounted for 51% of the total number, male accounted for 47% of the total number, and the rest were transexual.

Pie Chart of Gender

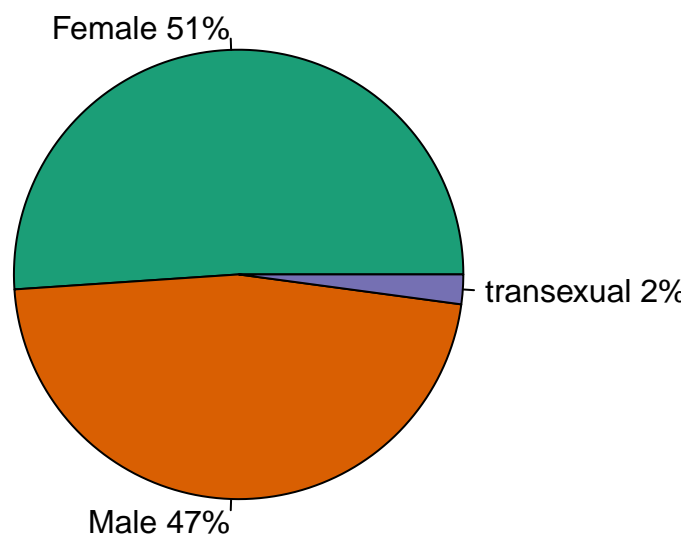


Figure 1: Gender ratio chart

2.4.2 Mean time by test type

Table 1 shows the average time to complete a Trail Making Test with or without background music. The results in Table 1 show that, compared with no music, it takes longer for students to complete the test with background music. The average time to complete a test with background music is 47.19 seconds(SD= 13.53), while the average time to complete a test in silence is 44.26 seconds(SD= 13.30), which is approximate 3 seconds faster than with background music.

Table 1: Average time to complete a test with and without background music

Test_Type	Average Time to Complete a Test	Standard Deviation
Music	47.19	13.53
No Music	44.26	13.30

2.4.3 Average time of completing a test by type and order

Since the experiment is a within-subject experimental design, each student is both in a control group and a treatment group. In order to study whether the practice effect exists, Table 2 shows that in addition to studying the average time for students to complete the test in each experiment condition, the factor of the order of completing the test is also added. Overall, regardless of whether students completed the first round of the test with background music or quietly, the average completion time of the second round was shorter than the average time of the first round. Table 2 shows that when students' first round is added intervention, the average time to complete the test is around 52.56 seconds(SD= 12.14). The second round, which is without background music, takes around 40.56 seconds(SD= 15.52) to accomplish the test. For the students who finish the test without background music in the first round, the average time to complete the test is 49.25 seconds(SD= 12.06). When added intervention, the background music, to the second round, the average time for them to complete the test not only did not increase, but it decreased by one second, about 39.95 seconds(SD= 10.18).

Table 2: Average time of completing a test by type and order

order	Test_Type	Average Time to Complete a Test	Standard Deviation
1	Music	52.56	12.14
1	No Music	49.25	15.52
2	Music	39.95	12.06
2	No Music	40.56	10.18

2.4.4 Average time for different genders to complete the test by test type

The paper(Goldman 2017) written by Bruce Goldman illustrates that females are good at concentration and not easily being distracted compared with males. In order to study whether background music will affect students' attention levels differently due to different genders, Figure 2 is drawn. Overall, the conclusion from the graph goes along with the conclusion Bruce Goldman made(Goldman 2017): females are less affected by the intervention. At the same time, males' attention levels are easily affected. The descriptive analysis result shows that background music increased the average time for female and transexual students to complete the test by about one second. In contrast, background music influences male students' average time to complete the test increased by 5 seconds. Moreover, whether with or without background music, the average completion time for men is the longest, while the average time for transexuals is the shortest.

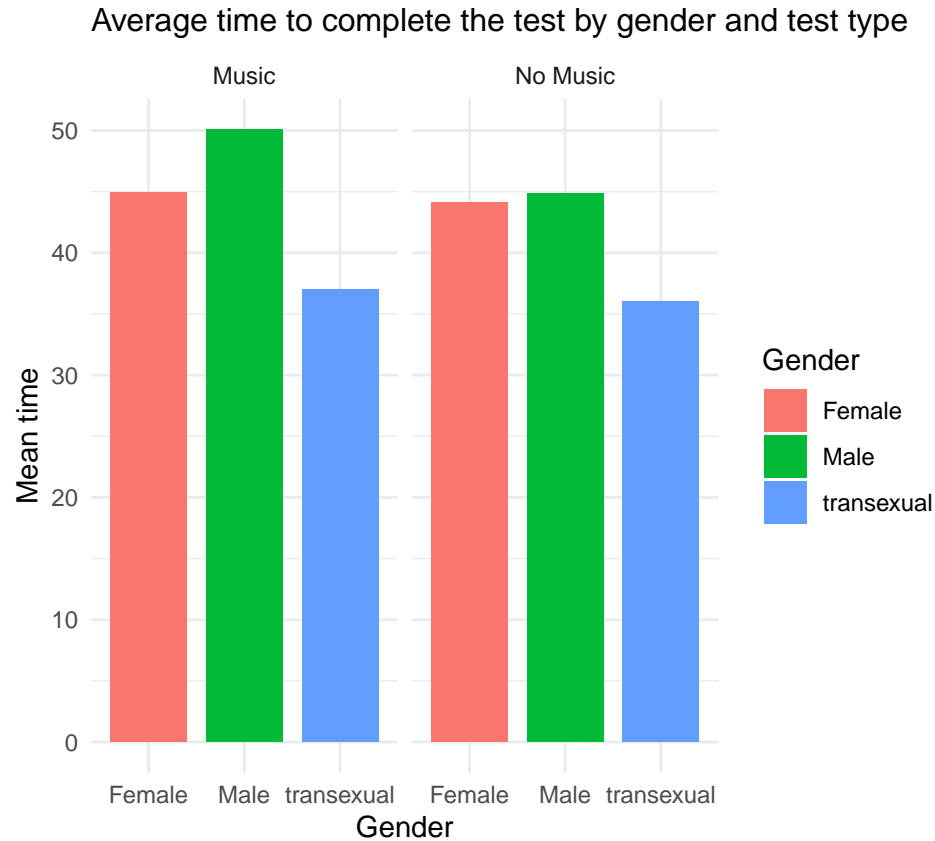


Figure 2: Average time for different genders to complete the test by test type

2.4.5 Music favorability and average time to complete the test

The line chart shown in Figure 3 is used to visualize the relationship between the degree of enjoyment of music from students and their attention levels. The blue line indicates the average time to complete the test at each rating level, and the brown dash line represents the total average time for accomplishing the test. As we can see from the chart, when the rating is approximately less than 3, the average time for students to complete the test is all greater than the overall average. When the rating is greater than 3, the average time for all students is significantly less than the overall average except for level 4.0. According to the findings in Figure 3, when students listen to songs they do not like while studying or doing other things that require concentration, their attention level is easily affected.

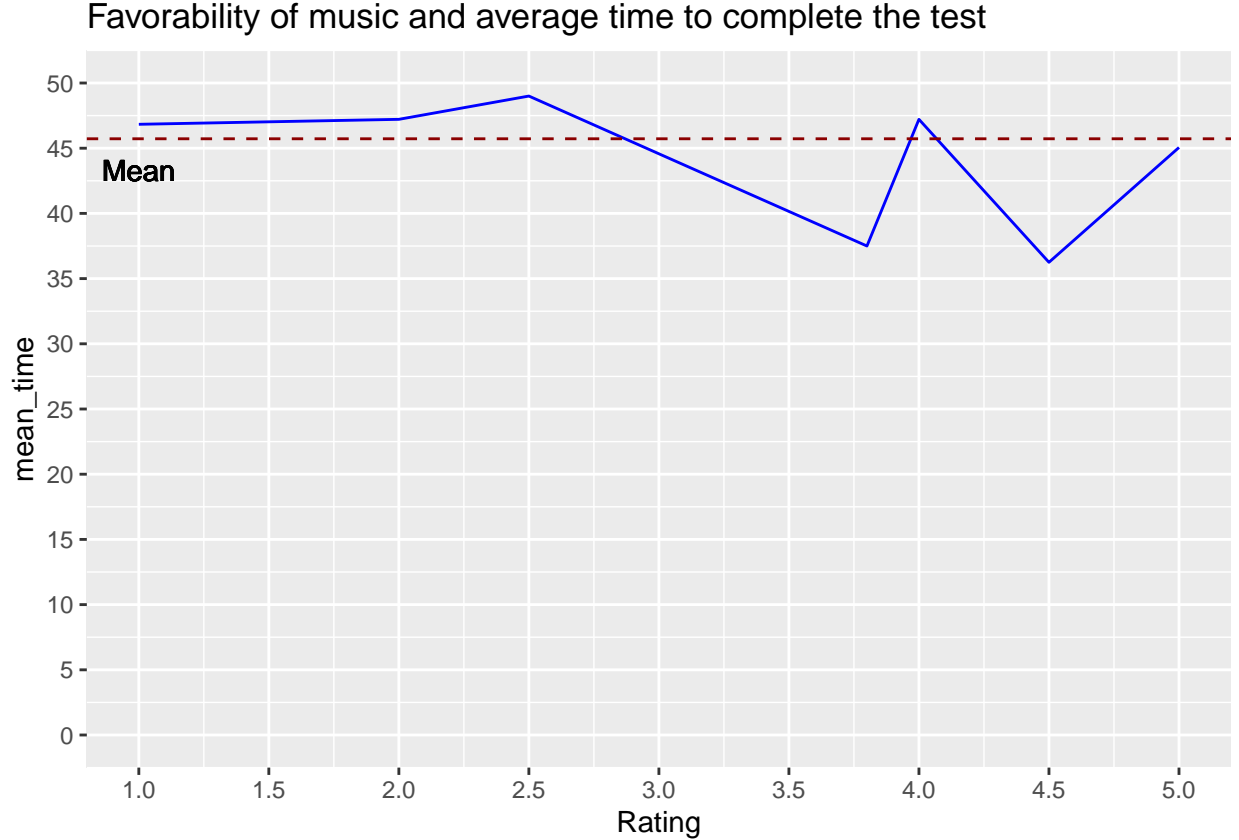


Figure 3: Music favorability and average time to complete the test

3 Model

The dataset collected from the within-subject experimental design mentioned in the paper is longitudinal data, which means one participant provides multiple data points⁵. These data will be related to each other because they are from the same participant. In this case, the basic linear regression model cannot be applied as it assumes that data points are independent of one another. Therefore, a linear mixed model provides an elegant model to handle the correlated observations within a group and suitable for the study with multiple trials per participant.

The analysis in this paper will be done entirely through R(R Core Team 2020). Figure 4 below shows the completion time of each student in a different order and test environment. One line represents two Trail Making Tests completed by a student, and the colours help distinguish the test environment. For example, a red line indicates that the student completed the first round of the test with background music and then completed the second round of the test in a quiet environment. Regardless of the experimental environment in which the first round is conducted, Figure 4 shows that different people have different initial attention levels, which indicates that there are random effects, predominantly random intercepts. In addition, the slope for each line is different, which imply that random slope also exist. So, the linear mixed model is applicable. It quantified how background music played in the background could impact the students' attention level.

⁵In the experiment, each student made the Trail Making test twice, one with background music, the other one without background music.

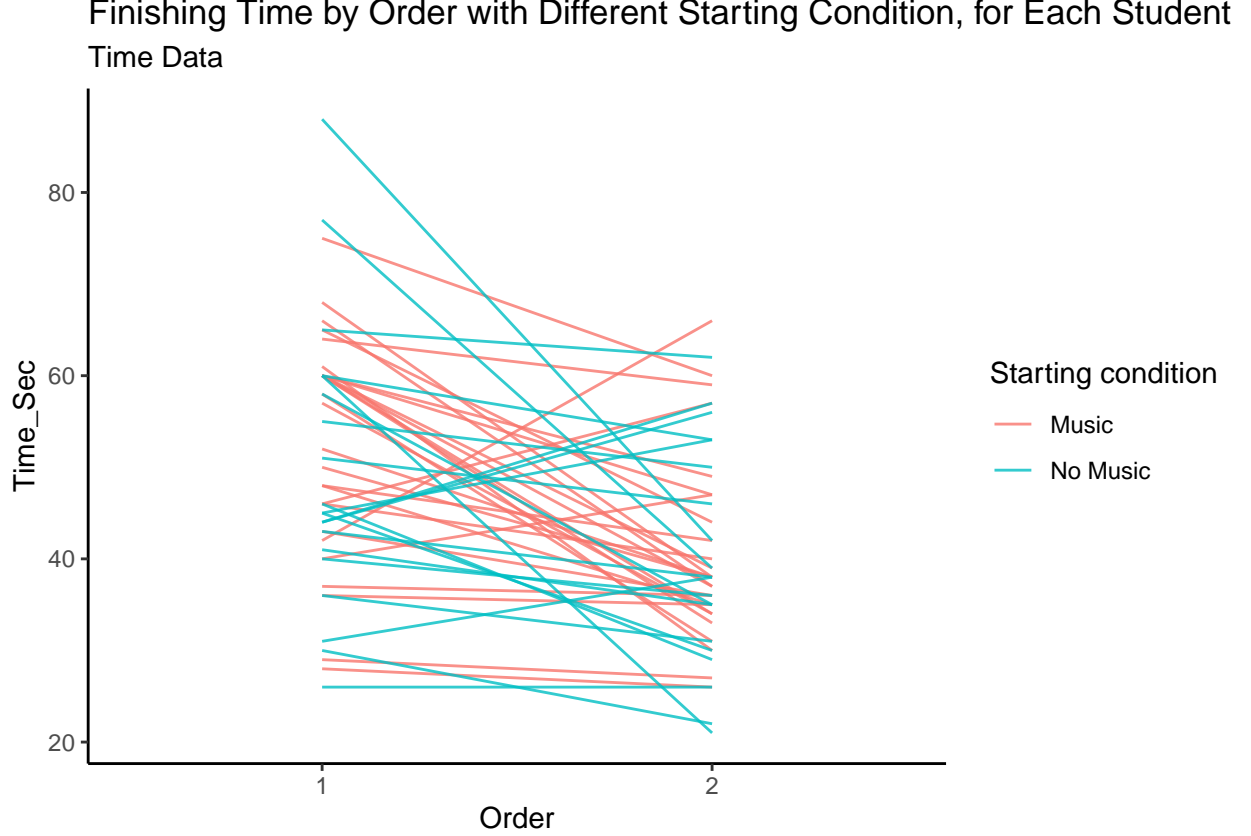


Figure 4: Finishing Time by Order with Different Starting Condition, for Each Student

3.1 Linear Mixed Model

$$\hat{Y}_{TimeSec} = \hat{\beta}_0 + \hat{\beta}_1 * I_{No\ Music_{ij}} + \sum_{a=Male}^{Transsexual} \hat{\beta}_2 * Gender_{aij} + \hat{\beta}_3 * I_{order=2ij} + \hat{\beta}_4 * Rating_{ij} + \alpha_{ij} + \epsilon_{ij}$$

Time is the time of the i -th students at the j round to complete the Trail Making Test. **The beta** shown in the formula are the overall fixed intercept and fixed-effect coefficients. **The alpha** is a vector of the random effect for group identity, which includes random intercept and random slope effect for the i -th student. **Epsilon** is the random error for the observation j in group i . **The variables** are the fixed-effect regressors for observation j in group i .

According to the dataset, the response variable and the explanatory variable, Rating, are continuous variables in the model. The reason is that they all can take any value between their minimum value and their maximum value (contributors 2021a). The other variables, type of test, gender and order, are all categorical variables because all of the data within the variables can be stored into groups or categories with names or labels (Blog 2019).

3.2 Modeling Procedure

A linear mixed model consisted of two-part, fixed effect and random effect. The modeling procedure will start to specify the random effects term that best fits the data in the null model, then building up hierarchically. ANOVA, p-value, AIC and RMSE help to find the best-fitted model for the data and check the model validation.

3.2.1 Random Effects

Three null models with different random effect terms will be established. The models will be compared using the ANOVA function to compare the p-values to see if the model adaptability is improved by adding other items.

The three different random effect terms are as follows:

- (1| Id): Random intercepts and slopes for each Id.
- (1 + Test_Type | Id): The effect of the background music will vary between each student Id. Random intercepts for test type, random slopes for students influenced by test type.
- (1 + Test_Type * order | Id): Each student Id can have their intercept, random slopes influenced by Test Type and order, and their interaction between Test Type and order.

3.2.2 Fixed Effects

After finding the best-fitting random effect term, ANOVA will be applied to select the significant variables and compare different models. The model for the study starts with only focused variable, test type. Then, other variables will be included in the model one at a time. From the ANOVA output, if the p-value were smaller than 0.1, the variable would be added to the final model. In the process of determining the best model, ANOVA will be applied and the smaller AIC⁶ suggests the better model.

3.3 Model Diagnostics

Before conducting the linear mixed model, the following assumptions need to be checked: whether the variance of the residuals is equal across groups, whether the residuals of the model are normally distributed, and whether there does not exist multicollinearity.

Figure 5 helps to check the assumption of the variance of the residuals is equal across groups. There seems to be an evenly spread around the central line from the chart(Figure 5) and no patterns evident, which means the assumption of homogeneity of variance being satisfied.

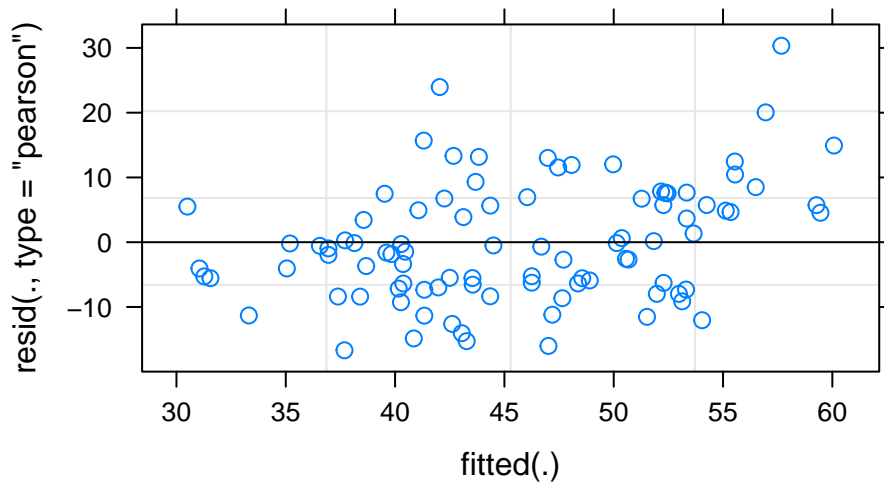


Figure 5: The variance of the residuals is equal across groups

⁶AIC(Akaike information criterion) is a mathematical method for evaluating how well a model fits the data generated. AIC is used to compare different models and determine which one is the best fit for the data(Bevans 2020)

QQ plot(Figure 6) can provide an estimation of where the standardized residuals lie for normal quantiles. Substantial deviation from the provided line indicates that the residuals themselves are not normally distributed. From Figure 6, there is some deviation from the expected normal line towards the two-side tails. However, overall, the line looks straight and, therefore, pretty normal and suggests that the assumption is not violated.

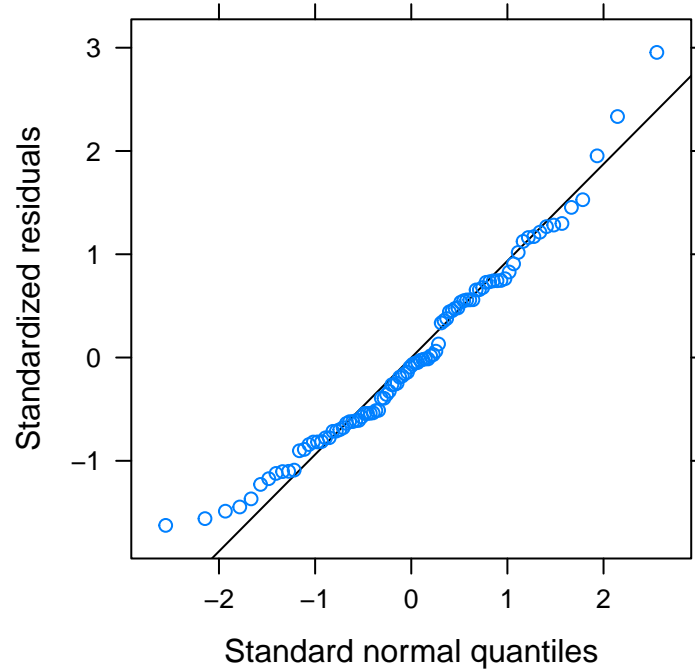


Figure 6: The residuals of the model are normally distributed

Multicollinearity means one independent variable can be linearly related to the others(contributors 2021b). Variance inflation factor(VIF) is utilized to verify the assumption of no multicollinearity because VIF is a measure of the amount of multicollinearity. The bar chart in Figure 7 illustrates that all the independent variables in the model have a VIF value less than 2.5, which means there are low, and even no multicollinearity exists.

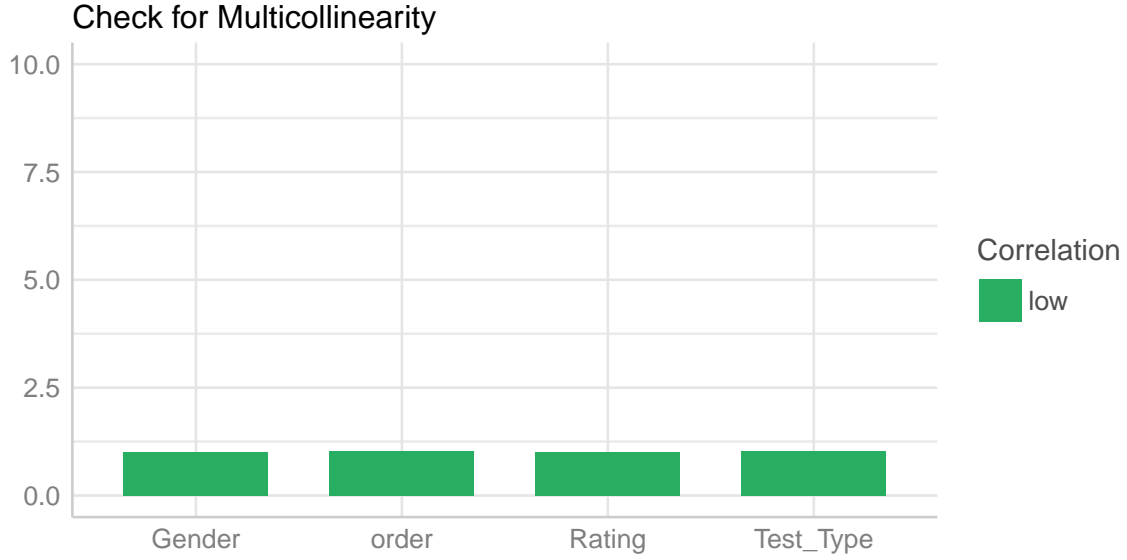


Figure 7: multicollinearity

4 Results

The purpose of the study was to explore whether background music has a distracting effect on students' attention levels. There were a total of 47 students participated in the study. A linear mixed model was used to analyze the results for the research question for the within-subject design. An alpha level of 0.1 was used for the analysis.

4.1 Random Effect Result

The output table(Table 3) using ANOVA to compare null models with different random effect terms shows that nullmodel1 is the best because the p-value for comparing nullmodel1 and nullmodel2 is larger than 0.1. There is no improvement from nullmodel2 to nullmodel3 due to a high p-value,0.994. As a result, the best random effect term is (1|Id).

Table 3: The result of comparing model by ANOVA

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
nullmodel1	3	759.79	767.42	-376.89	753.79	NA	NA	NA
nullmodel2	5	761.26	773.98	-375.63	751.26	2.53	2	0.28
nullmodel3	12	774.22	804.74	-375.11	750.22	1.04	7	0.99

4.2 Full Model Result

After finding the best-fit random effect term, the predictor variables will be added to the model one by one. Model 1 in the table only includes the most important variable, Test_Type. The result in the table(Table 4) shows that the coefficient of Test_Type is -2.936 with a 90% confidence level is -8.064 to 2.192, which illustrates that the average time for completing the Trail Making Test with or without background music is not significantly different between groups. Then, the variable order was added to model 2. The result from Table 4 shows that the coefficient for the variable order is -10.65 and the 90% confidence interval is

-14.848 to -6.452, which does not cover value 0, suggesting that the average time to complete the test in order 2 is 10.65 seconds faster than the order 1, holding other variables constant. Meanwhile, for model 3, the interaction between test type and order being considered for testing whether the effect of the order variable depending on the level of test type. The output from model 3 shows that the interaction term is not significant, p-value larger than 0.1 and the 90% confidence interval cover value 0. Furthermore, variable Rating was added to model 4 and showed not significantly related to completing the test. Finally, the same logic applied to model 5, which plus variable Gender and illustrate that the different Gender would not be a factor to influence student's time of accomplishing the test.

In order to find out the best-fit model for the data, ANOVA and AIC were applied to compare models. In Table 4, there is an AIC value in each model. The smaller the AIC value, the better the model. From Model 1 to Model 5, their AIC values are 760.54, 742.68, 744.21, 746.02, and 747.68. Among them, the AIC value of Model 2 is the smallest. Therefore, it is deduced that Model 2 should be the best model among the five models. To further prove that model two is the best model among the five, ANOVA is used to compare the five models. Table 5 analyzes which model is the best by looking at the p-value. When comparing model 1 and model 2, the p-value is much smaller than 0.01, which shows that model two is significantly better than model one, so model two is the best at this time. When comparing model two and model three, the p-value is 0.497, which is greater than 0.1, so adding predictor variables does not increase the fitness of the model. The same logic applies to the comparison between Model 3 and Model 4 and Model 4 and Model 5. In conclusion, through ANOVA's analysis, the conclusion is that Model 2 is still the best.

Table 4: Linear Mixed Model Result

	<i>Dependent variable:</i>				
	Time_Sec				
	(1)	(2)	(3)	(4)	(5)
Test_TypeNo Music	-2.936 (2.616)	-1.350 (2.142)	-3.306 (3.584)	-2.992 (3.650)	-3.891 (3.638)
order2		-10.650*** (2.142)	-12.606*** (3.584)	-12.292*** (3.650)	-13.191*** (3.638)
Rating				-0.489 (1.115)	-0.316 (1.100)
GenderMale					3.155 (2.853)
Gendertransexual					-9.132 (9.738)
Test_TypeNo Music:order2			3.911 (5.746)	3.285 (5.910)	5.083 (5.882)
Constant	47.191*** (1.936)	51.723*** (1.997)	52.556*** (2.338)	54.058*** (4.147)	52.579*** (4.421)
Observations	94	94	94	94	94
Log Likelihood	-376.271	-366.338	-366.107	-366.011	-364.841
Akaike Inf. Crit.	760.542	742.675	744.214	746.022	747.682
Bayesian Inf. Crit.	770.716	755.392	759.474	763.826	770.572

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 5: To Find Out the Best Final Model by ANOVA

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
m1	4	760.54	770.72	-376.27	752.54	NA	NA	NA
m2	5	742.68	755.39	-366.34	732.68	19.87	1	0.00
m3	6	744.21	759.47	-366.11	732.21	0.46	1	0.50
m4	7	746.02	763.83	-366.01	732.02	0.19	1	0.66
m5	9	747.68	770.57	-364.84	729.68	2.34	2	0.31

4.3 Model Assessment Result

Root Mean Square Error(RMSE) is used to evaluate the quality of the model. RMSE stands for the standard deviation of the residuals and it helps to tell how concentrated the data is around the best fit line(To 2021). In Table 6, the RMSE value of the final model, model 2, is 0.135 which a good sign for the model and illustrates that the model has a good ability to predict the data accurately. The R-square is applied to assess the model as well. R-square is a statistical measure of how close the data are to the fitted regression line(Blog 2013). R-square consisted two types, which are marginal R-square and conditional R-square for the linear mixed model. The paper will mainly focused on marginal R-square values, which provides the variance explained by only fixed effect. The marginal R-square for the final model, model 2, shown in Table 6 is 0.169, which means the final linear mixed model can explain 16.9% of the variation of the dependent variable. From the statistical result of the final model shown in appendix , the random effect result illustrates that the variance of Id is 43.08 and the variance of residual is 105.41. The variability due to the Id is 29%, which is means 29% within treatment variance is explained by individual differences. It is suitable to use a linear mixed model for this data.

Table 6: Assessment Result

RMSE	Marginal R-square	Conditional R-square	Variance of Id
0.134	0.169	0.41	0.29

5 Discussion

5.1 Ethics statement

The experiment conducted in this paper is about learning whether background music will influence students' attention level. The intervention in the study is background music. There is no potential to harm or endanger any participant. Moreover, the designer did not collect any sensitive data. Therefore, no formal ethics approval is required.

5.2 Descriptive Analysis Results

5.2.1 Different order will lead to different results

The data is divided into two main parts according to whether there are interference items. Then, add subgroup according to order in each group. A part of the students is randomly selected to complete the test with background music and then complete the test without background music. The other students oppositely first completed the test without background music and then completed the test with background music. The results showed that, overall, students generally took longer to complete the test with background music. The result of order illustrates that whether it is background music or not, the completion time of

the second round is less than the completion time of the first round. Therefore, it can be predicted that the order will affect the student's completion time.

5.2.2 Male are more susceptible to music interference

When comparing the average time to complete the test with and without background music for different genders, the results show that the change in female and transexual people is not apparent. After adding interference factors, the average time to complete the test increases by one second. However, the average time for males to complete the test was 45 seconds before adding the interference factor. After the interference factor was added, the average time to complete the test was 50 seconds. The background music increased the completion time for the male by 5 seconds. Therefore, in all the participants in this experiment, males are more likely to be disturbed by background music, which affects their attention level.

5.2.3 The enjoyment of music will affect the completion time of the test

After completing two rounds of testing, participants were asked to rate the background music, 1 to 5 anonymously. The higher the score, represents the more they liked the song. From the results, when the participants give a score greater than or equal to 3, the average time to complete the test is less than the overall average completion time. This shows that when people hear the background music they don't like, their attention level is much lower than those who listen to background music they like.

5.3 Model Results

5.3.1 Each student's initial attention level is different

When checking linear mixed model validation, the resulting plot shows that there exist random intercepts and random slopes, which illustrates that each students' initial attention level is different and fixed effects will vary for each student. In other words, the linear mixed model is suitable for this data and can incorporate student-to-student variability in finishing time and improve the ability to describe how fixed effects relate to outcomes.

5.3.2 Background music does not significantly affect the level of attention of students

Although in the descriptive analysis, it is found that in the case of background music, the average time for students to complete the test is longer than that without background music. But in the process of comparing the five models, the results show that `test_type` is not significant among all the models. This shows that it fails to prove that background music will affect the level of attention of students in a statistical sense.

5.3.3 The order of testing will affect the time to complete the test

When comparing models, the results show that the order significantly affects the time it takes for students to complete the test. In the final best-fit model, the coefficient of order is -10.65, and the 90% confidence interval is from -14.85 to -6.45. The coefficient, -10.65, means, on average, the time for completing the test in second-order is 10.65 seconds faster than the time in the first order when holding other variables constant.

5.4 Practical Implications

Although the model result of the study demonstrates that background music does not significantly impact students' attention level, it also not proved that background music fostered attention level. For students,

this is merely a matter of personal preference, such as whether they want to play background music while studying, for example, in an attempt to raise their motivation levels by playing background music.

Judging from the results of Ming Chou's(Chou 2010) study on which music types can distract students, hip-hop music is the most distracting, followed by classical music. The best performance is the quiet environment. According to the results of this paper, students should be careful when choosing which background music to listen to. Luckily, there is enough music readily available so that each of us has the chance to listen to our preferred music. Although the music rating is not significantly related to the completion time, the descriptive analysis shows that listening to the preferred music will improve attention level. The results may not be representative, but at least for the students in the experiment, listening to their favourite background music during the study may be conducive to attention level improvement and learning.

5.5 Weaknesses and Next Steps

5.5.1 Weaknesses

Practice Effect: Each participant is measured for two conditions in the within-subject design(repeated measure design). Participants learn about the type and difficulty of the question when they take the Trail Making Test for the first time, so they may become more adept simply with the question and more proficient because of the repeated test, making the second-round test result more optimistic. In addition to the students' understanding of the test questions, some participants may feel bored or have a negative attitude to repeat the same test, making the results inaccurate.

Hawthorn Effect: During the experiment, participants know that they are part of the experiment and observed. Their speed of doing the tests and the performance will be recorded, leading to the Hawthorne effect. Hawthorne effect stands for the tendency of some people to perform better when they know they are the participants in an experiment. The existence of the Hawthorne effect may cause students to change their behavior, such as trying to complete the test as soon as possible to achieve better performance, precisely because they know they are receiving research attention. Then, the results may be biased and inaccurate.

Sample Size: The sample size was the major limitation of this study. Large samples could have provided more reliable results. Small sample size may cause a large standard deviation, which would lead the result to be less accurate and decreasingly representative of the entire population. Future research should expand sample size.

Additional experimental materials: The study was conducted to test how background music affected participants' attention levels in two groups. In the within-subject design, the researcher would require to prepare two different Trail Making Test for each participant due to each student needs to test twice. For the control and treatment groups in the between-subject design, only one Trail Making Test was needed because each group only sees the material once. Therefore, using a within-subject design, the individual difference of participants is reduced but instead projects problems with individual differences between the materials participant are exposed to. The materials must therefore be carefully examined to ensure equal quality in factors such as difficulty.

5.5.2 Next Steps

Between-subject Design with different types of music: After expanding the sample size, change the experiment design to a between-subjects study experiment. Between-subjects design means each participant will only test each situation once, which prevents practice effect by eliminating the transfer of knowledge and shorten testing sessions by allowing each participant only need to do the test once.

In this paper, due to the small number of samples, only one type of background music is used for the test, namely popular music. Since the music used in the experiment in this paper is the most popular music at the time. However, given that some people may have heard it or are very familiar with it, and some may not have heard it, the impact of the song on the attention of each student may be different. In future research,

several types of music that are not familiar to everyone can be used for testing, for example, hip-hop music, traditional music, lyrical music, classical music, and music with lyrics.

Stroop Test: Use Stroop Test⁷ instead of Trail Making Test for the experiment to test student's cognitive ability with background music. The Stroop Test consisted of two parts, Stroop Color and Word Test. The participants require to identify the colour of the colour word. For example, the word green with the colour black, the tester should choose the answer black. The Stroop test could better reflect whether background music distracts student's attention level by displaying the speed of processing information.

⁷The Stroop Color and Word Test (SCWT) is a neuropsychological test extensively used to assess the ability to inhibit cognitive interference that occurs when the processing of a specific stimulus feature impedes the simultaneous processing of a second stimulus attribute, well-known as the Stroop Effect (Federica Scarpina 2017)

6 Appendix

6.1 Test 1

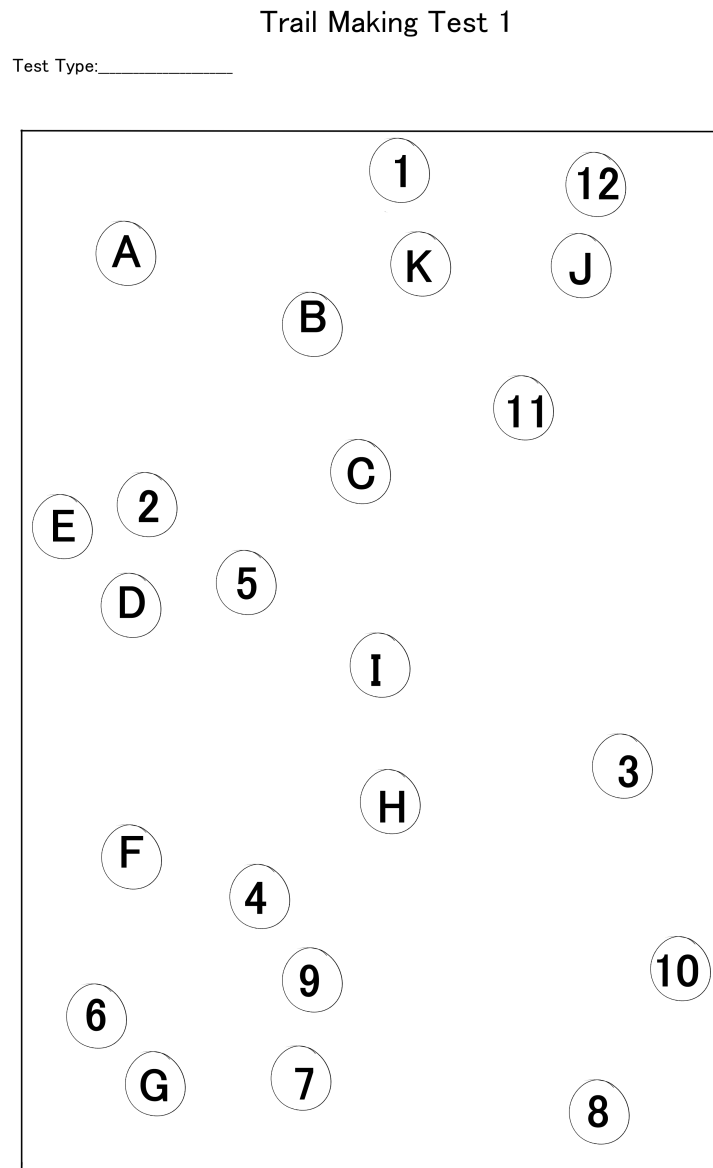


Figure 8: First Trail Making Test

6.2 Test 2

Trail Making Test 2

Test Type: _____

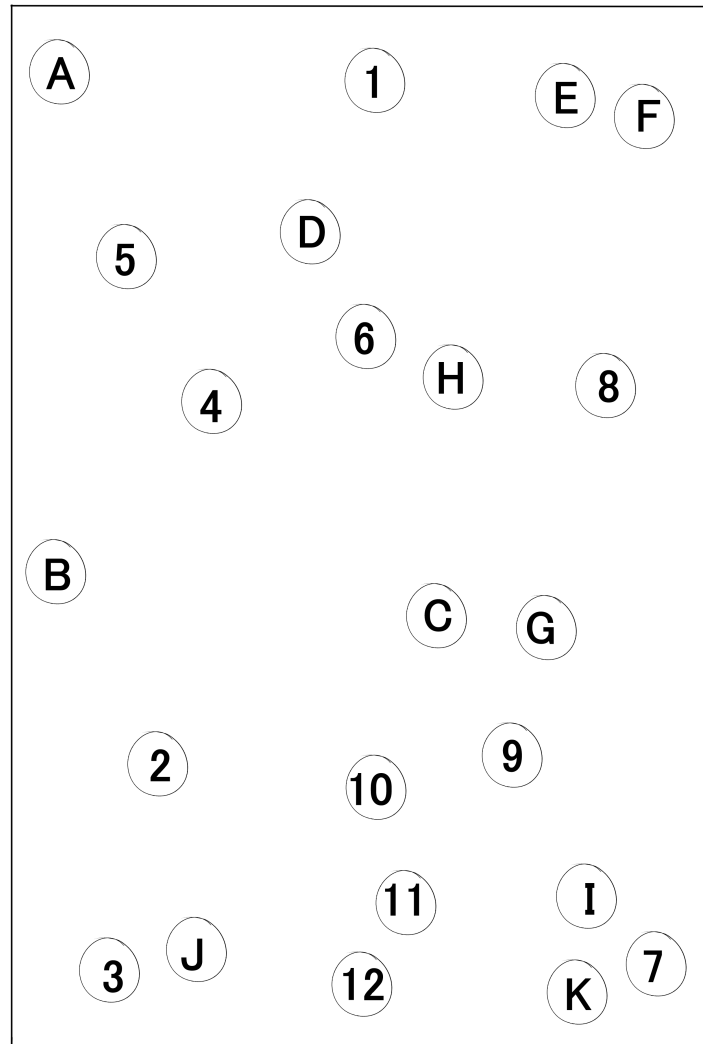


Figure 9: Second Trail Making Test

6.3 Survey

Elizabeth Blackburn School of Sciences
Extended Investigation Research Project

Post-experiment Survey

Please state your gender: (write)

Rate the music you listen to on a scale from 1-5: (circle)

Heavily Disliked 1 ----- 2 ----- 3 ----- 4 ----- 5 Heavily Enjoyed

Any Feedback for the experiment: (optional)

Figure 10: Short survey for rating

6.4 Final Model Output

```
Linear mixed model fit by maximum likelihood ['lmerMod']
Formula: Time_Sec ~ 1 + Test_Type + order + (1 | Id)
Data: dataset
```

AIC	BIC	logLik	deviance	df.resid
742.7	755.4	-366.3	732.7	89

Scaled residuals:

Min	1Q	Median	3Q	Max
-1.72635	-0.65659	-0.09236	0.64459	2.82040

Random effects:

Groups	Name	Variance	Std.Dev.
--------	------	----------	----------

Id (Intercept) 43.08 6.564
 Residual 105.41 10.267
 Number of obs: 94, groups: Id, 47

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	51.723	1.997	25.894
Test_TypeNo Music	-1.350	2.142	-0.630
order2	-10.650	2.142	-4.973

Correlation of Fixed Effects:

	(Intr)	Ts_TNM
Tst_TypNMsc	-0.456	
order2	-0.456	-0.149

References

- Bartoń, Kamil. 2020. *MuMIn: Multi-Model Inference*. <https://CRAN.R-project.org/package=MuMIn>.
- Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. “Fitting Linear Mixed-Effects Models Using lme4.” *Journal of Statistical Software* 67 (1): 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Bevans, R. 2020. “An Introduction to the Akaike Information Criterion.” [https://www.scribbr.com/statistics/akaike-information-criterion/#:~:text=The%20Akaike%20information%20criterion%20\(AIC,best%20fit%20for%20the%20data](https://www.scribbr.com/statistics/akaike-information-criterion/#:~:text=The%20Akaike%20information%20criterion%20(AIC,best%20fit%20for%20the%20data).
- Blog, F. 2019. “Categorical Vs Numerical Data: 15 Key Differences & Similarities.” <https://www.formpl.us/blog/categorical-numerical-data>.
- Blog, Minitab. 2013. “Regression Analysis: How Do I Interpret R-Squared and Assess the Goodness-of-Fit?” <https://blog.minitab.com/en/adventures-in-statistics-2/regression-analysis-how-do-i-interpret-r-squared-and-assess-the-goodness-of-fit>.
- Chou, Ming. 2010. “Attention Drainage Effect: How Background Music Effects Concentration in Taiwanese College Students.” <https://files.eric.ed.gov/fulltext/EJ882124.pdf>.
- contributors, Wikipedia. 2021a. “Continuous or Discrete Variable.” https://en.wikipedia.org/wiki/Continuous_or_discrete_variable.
- . 2021b. “Multicollinearity.” <https://en.wikipedia.org/wiki/Multicollinearity>.
- Dolegui, Arielle S. 2013. “The Impact of Listening to Music on Cognitive Performance.” <http://www.inquiriesjournal.com/articles/1657/the-impact-of-listening-to-music-on-cognitive-performance>.
- Federica Scarpina, Sofia Tagini. 2017. “The Stroop Color and Word Test.” [https://www.frontiersin.org/articles/10.3389/fpsyg.2017.00557/full#:~:text=The%20Stroop%20Color%20and%20Word%20Test%20\(SCWT\)%20is%20a%20neuropsychological,known%20as%20the%20Stroop%20Effect](https://www.frontiersin.org/articles/10.3389/fpsyg.2017.00557/full#:~:text=The%20Stroop%20Color%20and%20Word%20Test%20(SCWT)%20is%20a%20neuropsychological,known%20as%20the%20Stroop%20Effect).
- Goldman, Bruce. 2017. “The Cognitive Differences Between Men and Women.” <https://stanmed.stanford.edu/2017spring/how-mens-and-womens-brains-are-different.html#:~:text=In%20adulthood%2C%20women%20remain%20more,tend%20not%20to%20be%20gigantic>.
- Gutknecht, Jackson. 2018. “In a Tiger Hi-Line Survey, over Half of Students Listen to Music When Studying.” <http://www.hiline.cfschools.org/2018/11/in-a-tiger-hi-line-survey-over-half-of-students-listen-to-music-when-studying/>.
- Hlavac, Marek. 2018. *Stargazer: Well-Formatted Regression and Summary Statistics Tables*. Bratislava, Slovakia: Central European Labour Studies Institute (CELSI). <https://CRAN.R-project.org/package=stargazer>.
- Janina A. M. Lehmann, Tina Seufert. 2017. “The Influence of Background Music on Learning in the Light of Different Theoretical Perspectives and the Role of Working Memory Capacity.” <https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01902/full>.
- Lüdtke, Daniel, Dominique Makowski, Philip Waggoner, and Indrajeet Patil. 2020. “Performance: Assessment of Regression Models Performance.” *CRAN*. <https://doi.org/10.5281/zenodo.3952174>.
- Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. <https://CRAN.R-project.org/package=here>.
- Nielsen. 2017. “TIME with Tunes: HOW Technology Is Driving Music Consumption.” <https://www.nielsen.com/us/en/insights/article/2017/time-with-tunes-how-technology-is-driving-music-consumption/>.
- Philip David Zelazo, Michael T. Willoughby, Clancy B. Blair. 2016. “Executive Function: Implications for Education.” Institute of Education Sciences. <https://doi.org/NCER%202017-2000>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

- Sarkar, Deepayan. 2008. *Lattice: Multivariate Data Visualization with R*. New York: Springer. <http://lmdvr.r-forge.r-project.org>.
- Team, The Understood. 2021. “What Is Executive Function?” <https://www.understood.org/en/learning-thinking-differences/child-learning-disabilities/executive-functioning-issues/what-is-executive-function>.
- To, Statistics How. 2021. “RMSE: Root Mean Square Error.” <https://www.statisticshowto.com/probability-and-statistics/regression-analysis/rmse-root-mean-square-error/>.
- Wickham, Hadley. 2011. “The Split-Apply-Combine Strategy for Data Analysis.” *Journal of Statistical Software* 40 (1): 1–29. <http://www.jstatsoft.org/v40/i01/>.
- . 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Worsley, Arriken. 2020. “OSF.” <https://osf.io/9xh8e/>.
- Zhu, Hao. 2021. *KableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.