A Randomized Block Study of the Effects of Classical Music on Short-Term Memory Performance in Elderly Populations

Name: Zane Ching, Yuyao Huang, Oliver Pan, Jiawen Yu, Joshua Zhang

Date: Sep 9, 2025

Contents

1. Abstract

Classical music often has a societal correlation with intelligence and sophistication, and it might have scientific evidence to back this up. Following Rauscher *et al.*'s conclusion of the "Mozart effect" on spatial temporal reasoning, this experiment attempts to better understand the relationship between classical music and memory in elderly participants. Using a matched pair experimental design in order to control for individual variation of participants, 100 participants over the age of 60 were randomly selected from the Island and asked to take a memory test. Secondly, all 100 participants were assigned to listen to 10 minutes of classical music and then immediately perform the same memory test again. Before and after scores were recorded for each participant and analyzed using a paired t-test. Secondly, a linear regression model was run which predicted difference in memory scores before and after listening to music using the factors of number of hours slept, sex, height, and others, including anxiety level and how often they listen to music. The t-test found that on average, listening to 10 minutes of classical music had a statistically significant positive increase in memory score when measured immediately after listening. The linear regression concluded that classical music is especially useful for improving male short-term memory and is also effective in reducing anxiety and helping participants achieve better memory.

2. Introduction

Music is a powerful tool that can not only serve to provide stimulation in a traditional listening sense but also deliver emotional changes, which may have an effect on memory. According to the National Institute of Health (NIH), these changes can be attributed to the impact of music on arousal and mood. Studies have shown that listening to background music can increase productivity and quality of tasks such as memory (Ferreri), IQ (Cockerton et al.), and verbal and visual processing speed (Angel et al.), among other academic and behavioral benchmarks. On the other hand, music can also confound one's ability to perform tasks due to its potential to disturb or distract the subject from the task at hand, sometimes referred to as the "cognitive capacity hypothesis." An interesting angle to the study of the effect of music listening on cognitive performance is using older populations as test subjects; generally speaking, natural aging has degenerative effects on cognitive functions such as memory, so an elderly population can effectively represent a population with a lower tolerance of disruption of memory.

Yet another facet to this question of musical enhancement is the type of music being listened to. The NIH reports that faster tempo and major mode music can affect one's arousal and mood in the positive direction, while the opposite can actually make one sadder. As assumed that arousal and mood positively influence memory, sad emotions should be correlated with poorer memory test performance. This idea, pulled from the Mozart effect, refers to a 1993 study by Rauscher *et al.* which claimed that listening to Mozart for 10 minutes was associated with an increase of 8 to 9 points in IQ measurement for subjects, lasting for 10 to 15 minutes after musical stimulation. This study has caused major controversy in its wake, with one of its counterarguments being that with increased IQ performance being caused by the positive arousal effects of listening to Mozart, this should only be observed in subjects that *enjoy* Mozart. Finally, Rauscher reiterates that her observed effect can only explain improvements in spatial temporal reasoning and cannot be generalized to a broader increase in intelligence. However, directly implicated in spatial and temporal cognitive ability are working memory and long term memory. For this study, we decided to focus on the working memory of subjects before and after listening to classical music for a specific period of time.

3. Methodology

3.1 Participants

The participants of this study will be all residents aged over 60 living on the island, because the study objective is the elderly. A complete list of the population has been collected one by one by clicking on each island and randomly clicking on households, selecting only those that have residents aged 60 and above. These participants were first asked to complete a memory test. Then, they were asked to listen to classical music for 10 minutes, followed by another memory test. Finally, we conducted a survey on personal characteristics such as number of hours slept, sex, height, and other factors, including anxiety level and how often they listen to music.

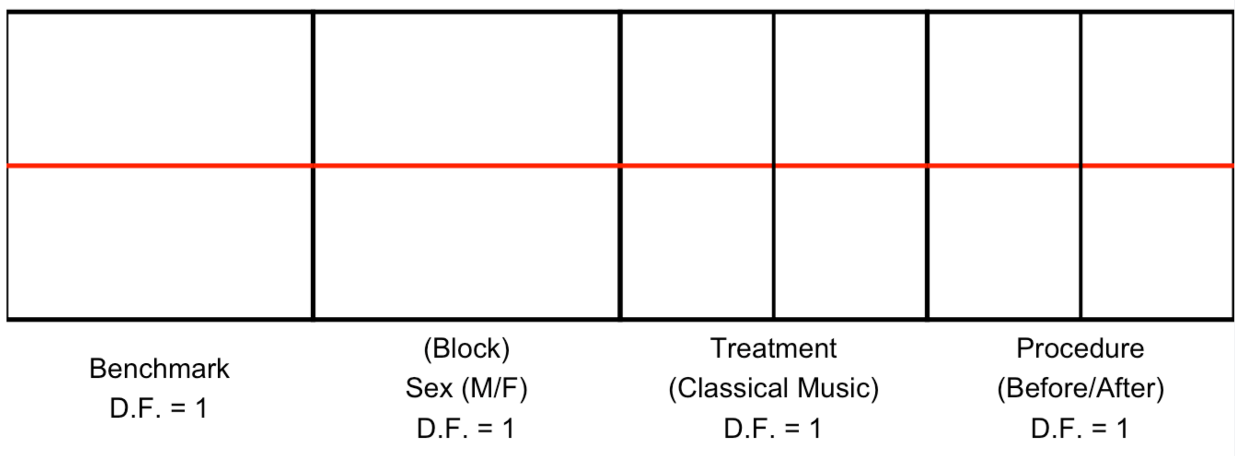
3.2 Design

The study will use a Matched Pair Design. The parameters for the design are summarized below:

Treatment Variable	Classical Music		
Procedure	Memory Test Before	Classical Music	Memory Test After
Response Variable	Memory Test After – Memory Test Before		

We chose to use a matched-pairs design because everyone has different reactions, body structures, and anxiety levels to classical music. A blocking design might not be able to divide the sample enough to compare how classical music affects memory. This approach is supported by prior quantitative research by Dionne Morris on The Effects of Classical Music on Memory. That paper indicated that groups performing memory tasks with classical music achieved significantly higher scores than those in non-music setting. However, our study improves on the experimental design by having each individual perform a memory test before and after listening to classical music, which helps eliminate possible confounding variables such as personal characteristics. A two-tailed t test in the earlier study showed a significant difference between the groups, indicating that classical music enhances memory. In this paper, we will also perform a t test, but specifically a paired t test.

The following is the factor diagram:



3.3 Instruments

Memory performance will be measured using a standardized recall test scored out of ten items. The test consists of presenting participants with a series of pokers for a fixed period of time, followed by a recall phase in which participants are asked to reproduce the series of pokers. This instrument was choose because it provides a direct and quantifiable measure of short term memory ability.

3.4 Procedure

Step 1: All residents aged 60 and above living on the island were considered, since the study objective focused on the elderly.

Step 2: A complete list of the population was compiled by clicking through each island and randomly clicking on households. Only households that included residents aged 60 and above were selected. This "random click" method served as the randomization procedure. Moreover, the chosen participants have the option to decline this experiment.

Step 3: Each participant was asked to complete a standardized memory test, with scores recorded as the Before Music measure.

Step 4: After the memory test, participants were asked to listen to 10 minutes of classical music to receive the treatment.

Step 5: After listening to classical music, participants completed the memory test again. Scores were recorded as the After Music measure.

Step 6: Each participant completed a short survey on demographics and lifestyle factors, including the number of hours they slept last time, sex, height, anxiety level, and depression level out of 10, and frequency of classical music listening. This survey is done after the memory test so the participants won't be too tired for the memory test.

4. Data Analysis

4.1 Type of Statistical Analysis

Using R, we will run different statistical tests on the data. Our main test is a paired t-test, because every subject completed the memory test before and after listening to classical music. This test will determine if the mean difference in scores is significantly different from zero. If the p-value is less than 0.05, this means classical music significantly improves memory test performance in the short term. We will also report a confidence interval and the effect size. As a robustness check, we will also conduct a Wilcoxon signed-rank test in case the differences are not normally distributed. This nonparametric test examines the median change and provides another way to ensure the result is consistent without requiring normality as an assumption.

Since we also collected background information (age, sleep hours, anxiety, energy, music rating), we will apply a linear model to the data. In this model, the memory test score difference is the main outcome, and sex is added as a block to see if men and women react differently to classical music, while the other variables are included as controls. This allows us to check if the effect of music remains after accounting for personal characteristics. Moreover, we can analyze whether the effects of classical music differ between males and females. We will check assumptions by examining QQ plots and running the Shapiro–Wilk test on the difference scores.

4.2 Sample Size determination

We decided to use a power of 0.80 on this analysis, which means that there is an 80% chance of correctly rejecting the null hypothesis when the null hypothesis is incorrect. The alpha level was set at 0.05, which represents the probability of rejecting the null hypothesis when it is actually true. For the effect size, we used a conservative estimate of 0.25.

The G*Power analysis is depicted in Figure 1. Based on the figure, the required total sample size is 101 participants in order to achieve the target power of 0.80. In our study, our sample size is 100, which is nearly identical to the required sample size. This means that our study is well-powered to detect the hypothesized effect size. Although we are one participant short of the exact calculated requirement, the actual achieved power with n = 100 is still effectively 0.80, which is sufficient for drawing meaningful statistical conclusions.

This G*Power analysis confirms that our design of using a matched-pairs t-test comparing memory test scores before and after listening to classical music has the appropriate balance of sample size, effect size, and significance level to test our hypothesis with adequate power.

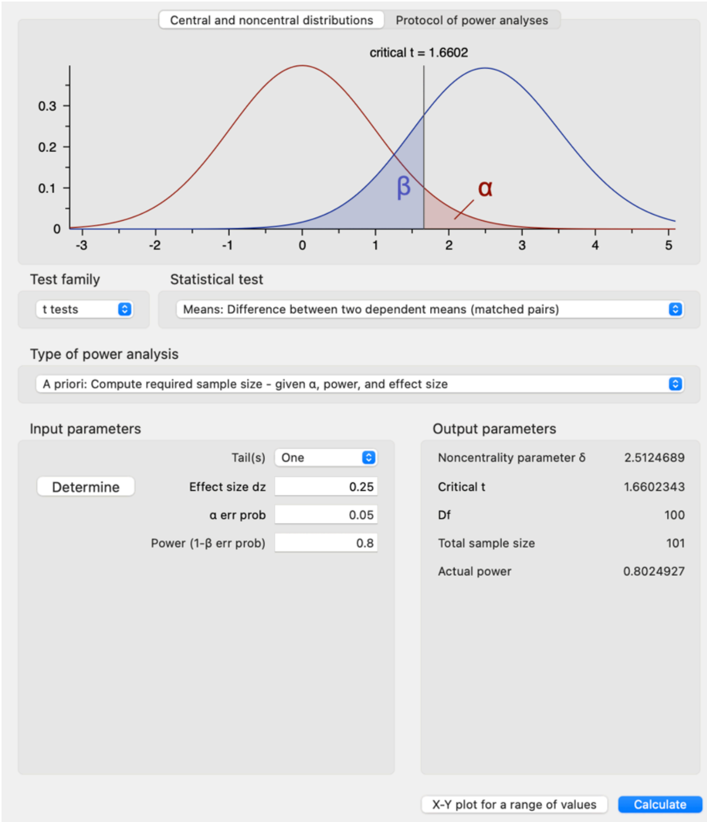


Figure 1

5. Results

5.1 Paired t-test Analysis

Test	t	df	p-value	Mean Difference	95% CI Low	95% CI High	Cohen's dz
Paired t-test	3.02	94	0.00326	0.442	0.151	0.733	0.31

Table 1: Results of Paired t-test Analysis on Memory Scores Before and After Listening to Classical Music

```
	Shapiro-Wilk normality test

data:  ds$Diff
W = 0.95038, p-value = 0.001243
```

Table 2: Results of Shapiro-Wilk Normality Test on Memory Score Differences

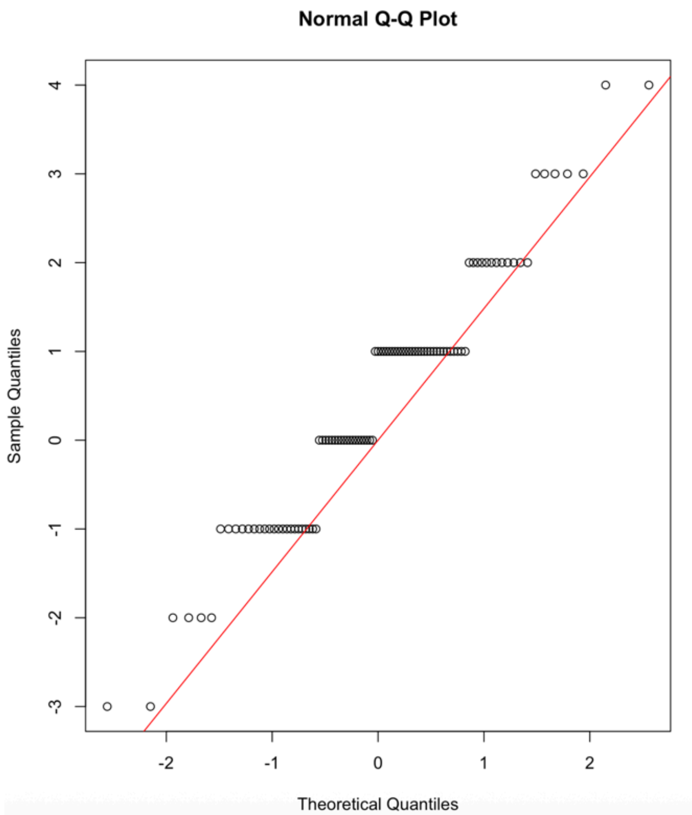


Figure 2: Normal QQ Plot of Memory Score Differences

To confirm our finding, we also performed a Wilcoxon signed-rank test, which does not assume normality of the difference scores. The Wilcoxon test also showed a significant improvement with a W statistic of 2000 and a p-value of 0.00385, which is less than 0.05. The Hodges-Lehmann estimate for the median difference was 0.5, with a 95% confidence interval between 0 and 1, which is greater than 0. This means the typical improvement in memory score was about half a point, and this result supports the previous conclusion above, that classical music has a positive effect on short-term memory performance. The fact that both the paired t-test and the Wilcoxon test gave consistent, significant results makes our conclusion more reliable.

Wilcoxon signed-rank test results for memory scores before vs. after classical music.

Test	W	p-value	Hodges–Lehmann Estimate	95% CI Low	95% CI High
Wilcoxon signed-rank	2000	0.00385	0.5	0	1

Table 3: Wilcoxon signed-rank test results for memory scores before and after classical music

5.2 Linear Model

Linear regression results for change in memory score (After – Before).

Predictor	Estimate	SE	t-value	p-value	95% CI Low	95% CI High
(Intercept)	4.382	3.372	1.30	0.1970	-2.318	11.083
SexFemale	-0.561	0.294	-1.91	0.0594	-1.144	0.023
Age	0.004	0.012	0.33	0.7440	-0.021	0.029
Hours	-0.562	0.397	-1.42	0.1600	-1.352	0.227
Energetic	-0.044	0.128	-0.35	0.7300	-0.300	0.211
Anxious	0.285	0.148	1.92	0.0579	-0.010	0.580
MusicRating	-0.055	0.076	-0.72	0.4740	-0.207	0.097

Table 4: Linear regression results for change in memory score (After-Before)

5.3 Residual Diagnostics

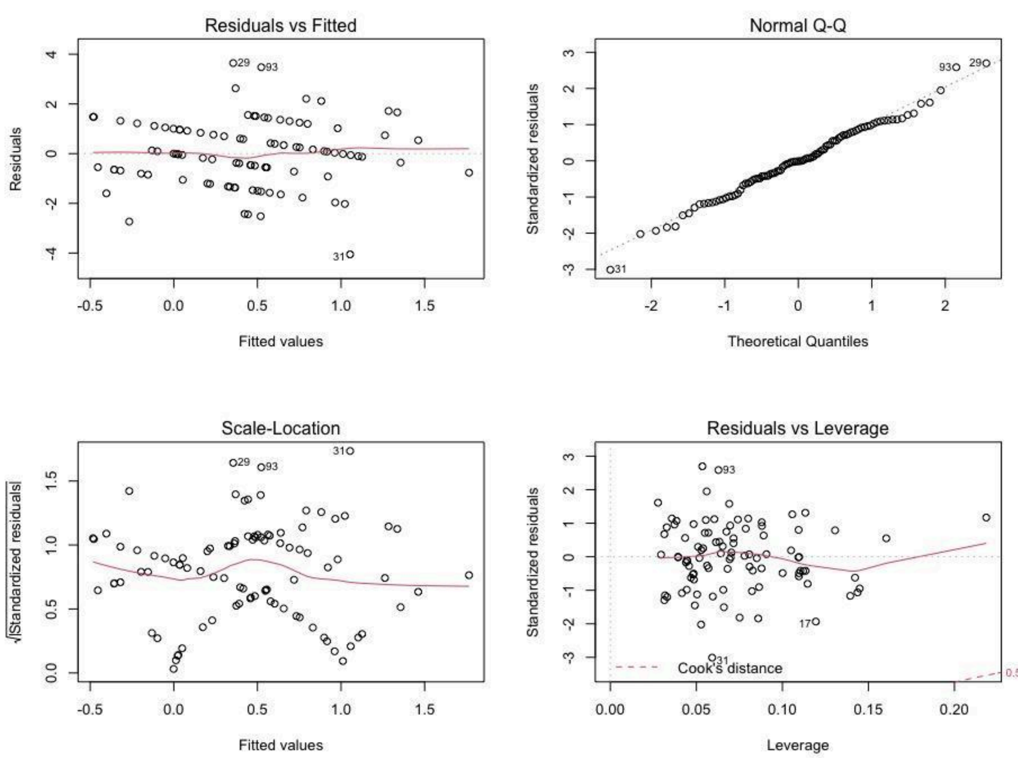


Figure 3: Summary Plots of Residuals for ANOVA Results.

5.4 Box Plots

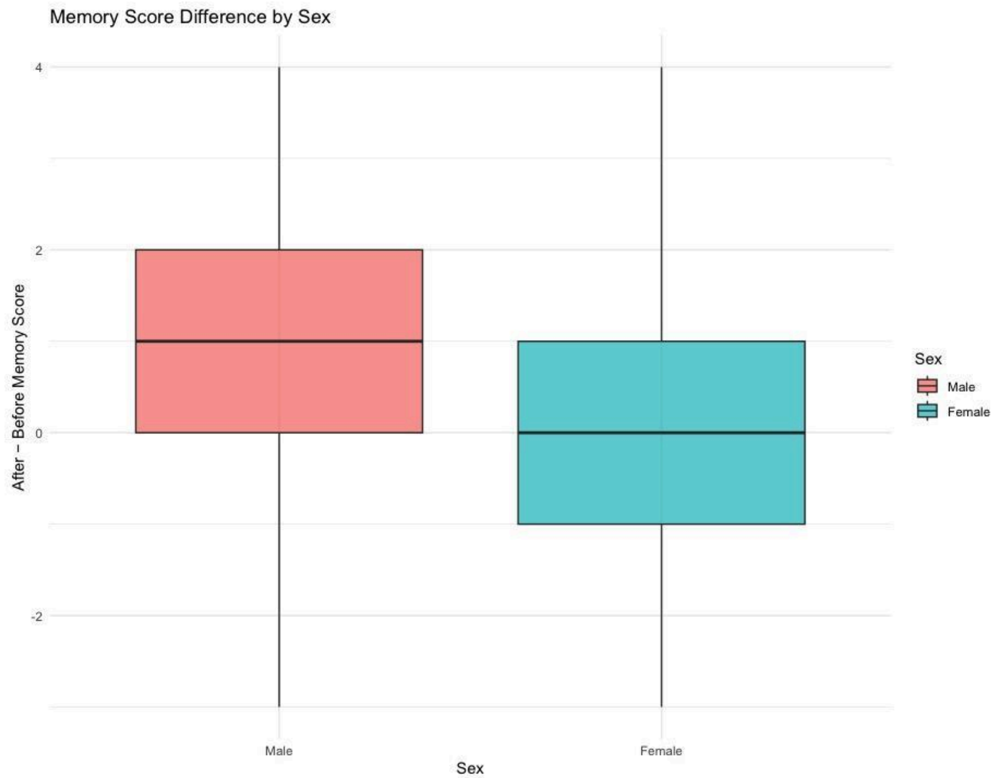


Figure 4: Box Plot of Memory Score Differences Before and After Listening to Classical Music by Sex

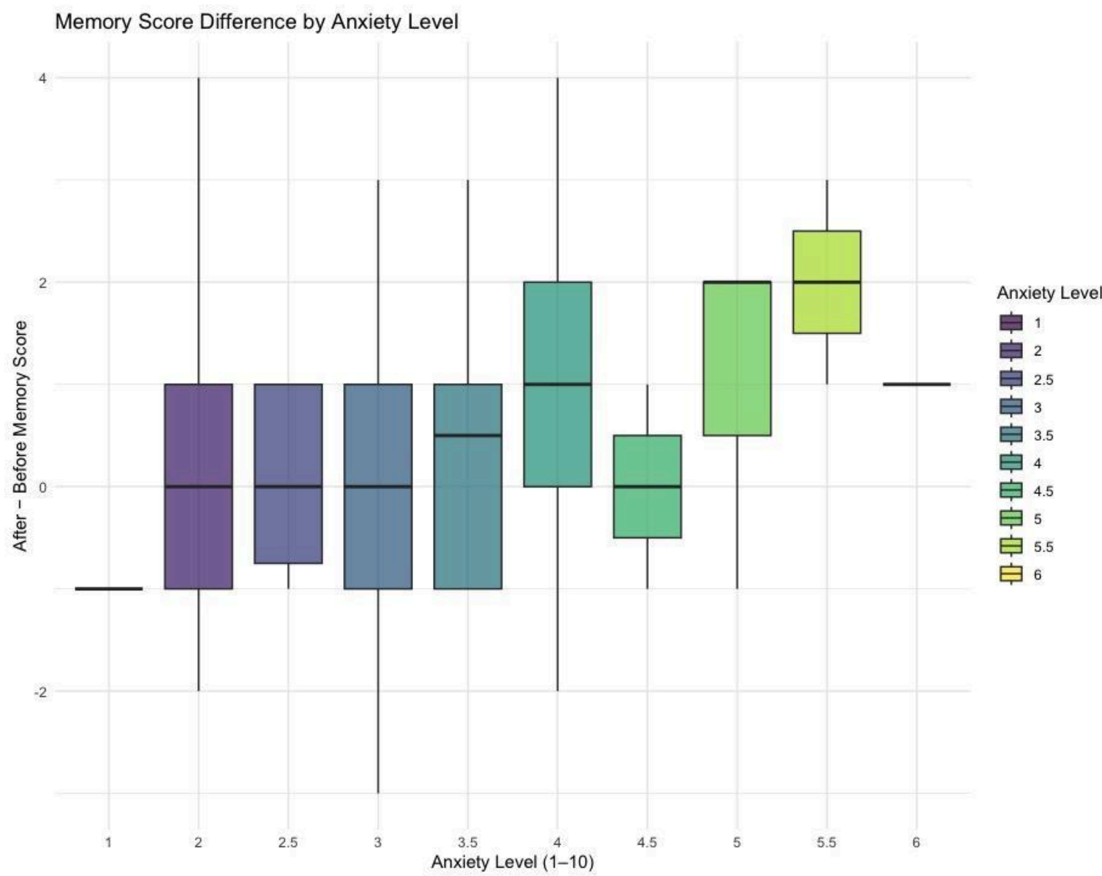


Figure 5: Box Plot of Memory Score Differences Before and After Listening to Classical Music by Anxiety Level

6. Discussion

The following is the result of the paired t-test, which showed that the average memory test score after listening to classical music was significantly higher than before, with a t value of 3.02, degrees of freedom of 94, and a p-value of 0.00326, which is less than 0.05. Therefore, we have significant evidence that the memory test score improved after the treatment. The mean difference was 0.442 points, with a 95% confidence interval from 0.151 to 0.733, which does not include 0 and also confirms our conclusion that there is a significant difference in memory test scores before and after the classical music treatment. The effect size, Cohen's dz = 0.31, suggests that on average, participants recalled almost half a point more items after the music, and since the confidence interval does not cross zero, we can conclude the improvement is statistically reliable.

In order to test the normal condition for the paired t test, the Shapiro–Wilk test for normality on the difference scores gave a W value of 0.950 and a p-value of 0.0012, which is less than 0.05. This result indicates that the distribution of the differences is not perfectly normal. The QQ plot also shows some deviations from the straight red line, especially at the tails, suggesting heavier tails and some outliers. Because the assumption of normality is not fully met, we also need to use the Wilcoxon signed-rank test as a nonparametric alternative to confirm our significant result.

In order to confirm that the difference in memory test scores is not only statistically greater than zero but also to quantify the effect among different genders and add in different controls, we also performed a linear model. The dependent variable in this linear model is the difference between after and before listening to classical music. Moreover, we want to see whether male and female participants have different effects on memory test scores after classical music.

The linear regression model result is shown below. The intercept was positive with an estimate of 4.38 but it was not statistically significant with a p value of 0.197. This means that after controlling for sex, age, sleep hours, energetic, anxious, and music rating, the overall improvement is still positive but not significant as desired. The coefficient for sexFemale was negative with an estimate of 0.561 and a p value of $0.0594 < 0.1$, which suggests that female participants improved slightly less than male participants, and this result is significant at the 10 percent level. This makes sense because female participants might already listen more to classical music on their own time, so the 10 minutes of classical music in the experiment might not be as effective for them as it is for male participants. The anxious score was positively related to the memory difference with an estimate of 0.285 and a p value of $0.0579 < 0.1$, meaning that participants who reported higher anxiety tended to show greater improvement after listening to classical music. This result is again significant at 10% significant level. Other variables, including age, hours slept, energy level, and music rating, were not significant. Overall, the regression confirms that the difference in memory scores remains even after controlling for personal characteristics, even if it is not statistically significant, and it shows two main conclusions. First, classical music is especially useful for improving male short-term memory, and second, it is effective in reducing anxiety and helping participants achieve better memory.

In order to confirm the robustness of the linear model done above, the diagnostic plots for the linear regression model are shown below. The residuals are spread around zero without a strong pattern in the Residual vs Fitted plot, although there are some outliers. The Normal Q-Q plot shows that most points follow the 45-degree straight line. Disregarding a few outliers in the Q-Q plot, this suggests the residuals are roughly normal and safe to perform the linear analysis. The scale-location plot shows that the spread of residuals is mostly even, but again a few outliers are present. The Residuals vs Leverage plot indicates that most observations have low leverage, with a few points having slightly higher influence but still within Cook's distance. Overall, the assumptions of linear regression are reasonably satisfied, although there are a few outliers and mild deviations from normality.

In order to visualize the differences we found in the linear model above, we have constructed a box plot for the memory score difference by sex. This box plot shows that male participants generally improved more than female participants after listening to classical music. The median difference for males is clearly above zero, while the median for females is close to zero. So, we can conclude the significant differences in the paired t-test result might be driven mainly by males. Overall, this visualization supports the regression result that male participants benefited more from the classical music treatment compared to female participants. This finding is interesting when compared to the study by Theofilidis et al., which showed that female college students achieved significantly higher short-term memory scores than males when exposed to different music genres, including classical music. While their study suggested a female advantage in memory recall under music conditions, our results indicate that in an elderly sample, male participants may benefit more from classical music exposure. This result shows that age and sample characteristics might influence the gender effect of music on memory (Theofilidis et al.).

In the second stage, we are trying to see the visualization of how memory score differs by anxiety level, so we constructed boxplots for different anxiety levels. The plot shows that participants with higher anxiety levels tended to benefit more from listening to classical music. At low anxiety levels, the median difference is close to zero, and this indicates that classical music has little to no effect on short-term memory. However, as anxiety levels increase, the median improvement rises above zero, and the interquartile range also shifts upward, showing that participants with higher anxiety generally have better short-term memory. This finding supports the idea that music does not just influence memory in isolation but also interacts with emotional states. As Zaatar et al. discuss in their paper "The transformative power of music: Insights into neuroplasticity, health, and disease", music engages a wide network of brain regions related to both cognition and emotion, and listening to music can help regulate brain activity and retrain circuits involved in stress and anxiety. This may explain why anxious participants in our sample gained more from the classical music treatment, as music provided an extra boost to both their emotional regulation and their memory performance.

7. Conclusion

With studies such as Rauscher *et al*'s "Mozart effect" raising the possibility that classical music might have a positive association with cognitive capacity or spatial temporal reasoning, we conducted this study to better understand how classical music affects the memory of elderly people. Using a matched-pairs experimental design on 100 participants randomly selected from the Island, we found a statistically significant increase in memory test scores immediately after listening to classical music when compared to memory test scores from pre-listening. The paired t-test and Wilcoxon signed-rank test confirmed this difference, with participants recalling nearly half a point more items on average.

With further analysis through a linear regression model, which analyzed difference in memory test scores pre- and post-listening with factors of number of hours slept, sex, height, and others, including anxiety level and how often they listen to music, it was found that listening to classical music for 10 minutes can especially work to improve male short-term memory over female short-term memory. Additionally, this regression found that classical music can be more beneficial to the memory of people who show higher anxiety levels. These findings suggest that classical music may not only enhance memory performance but also provide emotional regulation benefits that indirectly support cognition. While drug treatments and other cognitive interventions are often used to address memory decline in aging populations, our results highlight the potential of a simple, non-invasive treatment such as music listening.

This experiment only measured memory immediately after listening to classical music, so the duration of the effect remains unknown. Outliers in the data also suggest that individual differences in response to music may play a larger role than captured here. To strengthen our future research, we would recommend conducting studies with different age groups, extending time between memory tests or increasing the number of observations in a temporal manner, and comparing classical music with other genres. We might also conduct a blocked experiment where participants are divided into different treatments, such as a control and differing periods of classical music listening.

Overall, our findings support the conclusion that classical music can positively affect short-term memory in the elderly, particularly among men and individuals with higher anxiety. With many countries experiencing aging populations in the current time period, this research on memory is becoming increasingly more important. This work contributes to a growing literature on music and cognition and suggests that classical music may serve as a valuable tool in promoting mental well-being and cognitive health in aging populations.

8. Reference

Bottiroli S, Rosi A, Russo R, Vecchi T, Cavallini E. The cognitive effects of listening to background music on older adults: processing speed improves with upbeat music, while memory seems to benefit from both upbeat and downbeat music. *Frontiers in Aging Neuroscience*. 2014;6(284). doi:https://doi.org/10.3389/fnagi.2014.00284

Jenkins JS. The Mozart Effect. *Journal of the Royal Society of Medicine*. 2001;94(4):170-172. doi:https://doi.org/10.1177/014107680109400404

Mammarella N, Fairfield B, Cornoldi C. Does music enhance cognitive performance in healthy older adults? The Vivaldi effect. *Aging Clinical and Experimental Research*. 2007;19(5):394-399. doi:https://doi.org/10.1007/bf03324720

Morris, D. (n.d.). The effects of classical music on memory. Academia.edu. https://www.academia.edu/26790939/The_effects_of_classical_music_on_memory

Spencer, A. (n.d.). The effect of classical music on memory in the elderly. University of Queensland. https://islands.smp.uq.edu.au/academy.php?paper=music_memory

Theofilidis, A., Karakasi, M. V., Kevrekidis, D. P., Pavlidis, P., Sofologi, M., Trypsiannis, G., & Nimatoudis, J. (2020). Gender differences in short-term memory related to music genres. Progress in Neuro-Psychopharmacology & Biological Psychiatry, 104, 110021. https://doi.org/10.1016/j.pnpbp.2020.110021

UCLA Institute for Digital Research and Education. (n.d.). Power analysis for paired sample t-test | GPower data analysis examples*. UCLA: Statistical Consulting Group. https://stats.oarc.ucla.edu/other/gpower/power-analysis-for-paired-sample-t-test/

Zaatar, M. T., Alhakim, K., Enayeh, M., & Tamer, R. (2023). The transformative power of music: Insights into neuroplasticity, health, and disease. Frontiers in Neuroscience, 17, 10765015. https://pmc.ncbi.nlm.nih.gov/articles/PMC10765015/