

Comparative Cognitive Performance of Students in Different Environments.

Connor V. Hennen

1.0 Introduction and Statement of Problem:

Better understanding the conditions that lead to optimal and suboptimal cognitive performance would certainly be of great value in academic environments. Knowing the effects of happiness, stress, confidence, etc. on cognitive performance on cognitive ability could shape everything from the physical environment to the curricula in academia so as to foster optimal learning outcomes. The problem looked at here is limited to cognitive performance (specifically in arithmetic), not learning--for example, the speed at which a student might learn a new mathematical concept is not considered but the student's ability to express arithmetic skills in an examination is considered. However, the brain function for learning new concepts and performing old concepts probably aren't entirely unrelated, and performance alone is valuable and applicable to understand so the results of this study are still valuable in understanding and improving student learning outcomes.

Specifically, we consider three difference emotional and mental states, induced by external treatments, in this experiment. One state, induced by a Swedish massage, is marked by decreased stress and increased confidence. Another state, induced by eating 50 mg of milk chocolate, is marked by increased energy, slightly increased confidence, increased happiness, and decreased stress. The final state, induced by a lacerating the unfortunate student, is marked by anger, drastically decreased confidence, decreased happiness, and increased stress. Given that the sets of

emotions vary significantly, don't all address the same emotions, and don't isolate individual emotions, this study does not seek to identify which emotions improve cognitive performance and which worsen cognitive performance, but rather seeks to answer the question, "Does a difference exist between the performance of students under different emotional and mental states, induced by external conditions?" or, simply, "Does a student's mental and emotional state affect cognitive performance?"

2.0 Selection of Response Variable

The score of a student on a four minute arithmetic test is the selected response variable. This seemed to be a good choice for two main reasons. First, in the United States arithmetic, from addition to exponents, is covered in every K-3 curriculum, making the study a more realistic representation of actual academic conditions and routines. Second, arithmetic enables one to test cognitive performance and not cognitive knowledge, while partially eliminating the effects of experience and practice. This is a beneficial trait because it allows for a standard test to be applied to all ages in the study. For example, if the response variable was the score of a student on a calculus test then we could not reasonably apply it to five-year-old students. However, we can reasonably assume that the Island's four-minute arithmetic test does not contain anything that a five-year-old does not already know how to do and cannot be taught how to do. Thus, we can test all students from age five to twenty-three using the same test format.

3.0 Choice of Factors and Levels

A single factor of interest with three levels has been chosen and a single block factor with five levels has been chosen.

Factor of Interest: External Conditions Before Arithmetic Test

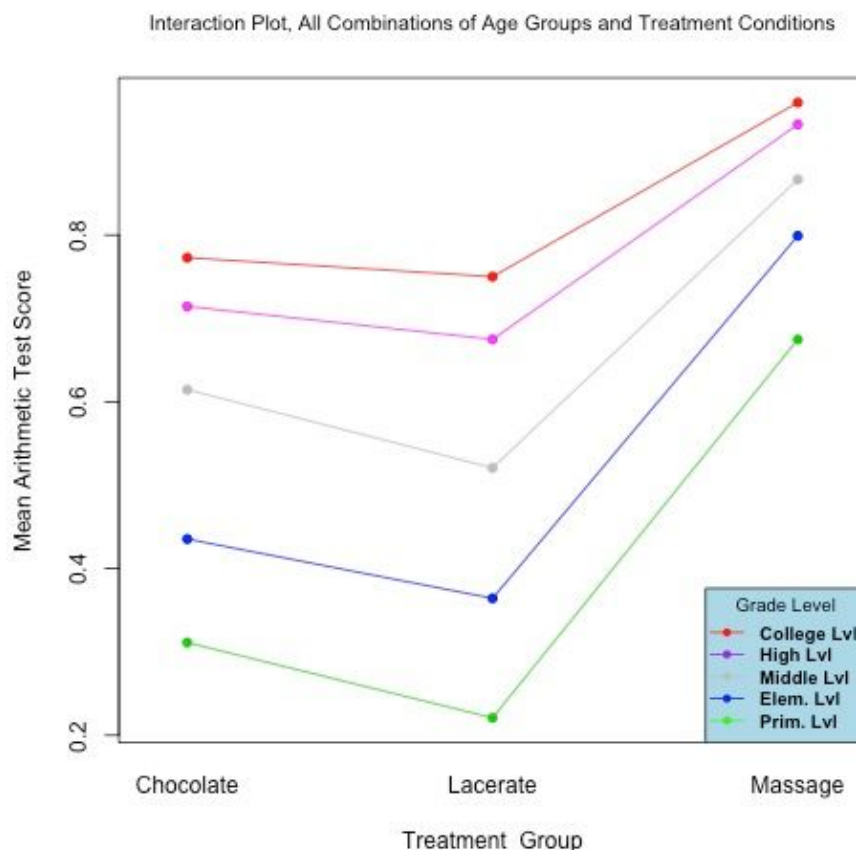
Three different conditions that would induce a varying spread of emotional and mental states were selected. As described in the introduction, the three levels are laceration, milk chocolate, and Swedish massage.

Block Factor: Age

Subjects were divided into five different age groups, 5-8 years old, 9-11 years old, 12-14 years old, 15-18 years old, and 19-23 years old. The groups were designed to mirror the different stages of academia (primary grade level, elementary grade level, middle school grade level, high school grade level, and college grade level) pretty closely while also maximizing the lowest number of subjects available for any group. The lowest number of subjects available for any group is relevant because in order to keep the experiment balanced I had to randomly select that number of subjects from each of the other age groups, discarding wasting those who did not get selected. The lowest number available for any group ended up being 600 so each block consisted of 600 subjects.

Blocking was a clear choice in this experiment because there would potentially be much higher variability in the mean aptitude of randomly selected groups. For example, though unlikely given the large sample size, it would be very misleading to treat a group of predominantly college students with a laceration and test them, and then treat a group a predominantly toddlers with a Swedish massage and test them, and then when

the mean score of the laceration group is much higher than that of the massage group, conclude that laceration improves performance. Blocking, however, makes sure that there are the same number of subjects per age groups for each treatment, as each age group block is divided evenly into the three treatments. Additionally different ages could handle stress differently and the blocks enable us to identify a noteworthy interaction effect as depicted below.



In general, as Islanders become older, they become more consistent and less vulnerable to external events and/or emotions. For example, for college age Islands scores fall very slightly when going from the chocolate group to the lacerate group, while for primary school level Islanders scores fall much more—by nearly 33.333%. This finding is actually valuable because it highlights the even greater importance of learning environments for lower grade levels.

FIGURE 1. Interaction Plot

4.0 Experiment Design

A randomized block design with one block factor and one factor of interest has been selected for this experiment. There will be 6000 total subjects, distributed evenly over all five age groups to render blocks of 600 subjects, which will then be randomly divided into three groups, one for each treatment, of 200 subjects. All in all, there will be 600 subjects per age group and 2000 subjects per treatment group.

The trials will be conducted using a Python bot that I built. The bot collects all of the village links from the Island homepage, then collects all of the hamlet links from the village homepages, then collects all of the house links from the hamlets pages, and finally all of the Islander links from the house pages. It then visits each Islander and asks for their consent to participate in the experiment, and, if consent is granted, adds the Islander to 'Contacts' and if consent is not granted, moves on.

Exploratory trials

In order to get a sense of which treatments invoked which emotions, I surveyed about Islanders before and after receiving a given treatment, asking how happy, angry, depressed, worthless, and energetic they felt. I tested many different treatments, from exercise to prescription drugs to being in freezing water

The findings were consistent for all of the Islanders so it did not require a formal experiment to conclude that the treatments each induced a certain set of emotions. For example, the results of the post-laceration surveys indicated increased anger, increased depression, decreased self-worth, and increased anxiety than did the results of the pre-laceration surveys for all 20 of the laceration subjects. Also, the magnitude of the differences in pre-treatment and post-treatment survey results were the same for all

subjects within each group except for a very small fraction--three or four out of twenty--of subjects in the groups which differed in the magnitude of difference by at most one point on a ten point scale.

Selecting Swedish massage, laceration, and milk chocolate as a set of treatments that would produce a wide range of different emotions, I automated the formal tests and web scraped the results with the Python bot. The testing procedure was just two steps: 1) Give the Islander the treatment they were randomly assigned to 2) Give the Islander a four minute arithmetic test. Because each block and each treatment group had such a large number of subjects, it was reasonable to assume that each of the treatment groups, because the subjects were randomly assigned treatments, would have essentially the same arithmetic test results as the other treatment groups within the same age group if all of the subjects were given the test prior to any experimental treatment. This is relevant because it allows us to compare treatment effects without having to compare pre-treatment and post-treatment scores for each subject. Instead we can just assume that the pre-treatment scores are the same as those of the other treatment groups within each block so we don't have to do a pre-treatment test and compare the mean change of the individuals.

5.0 Predictions

Before running the tests the final outcomes seemed fairly obvious with basic intuition. I predicted that, across all age groups, the massage treatment group would have the highest performance, followed by the chocolate treatment group, and then finally the laceration group would perform the worst. However I did not expect the mean

performance for the chocolate and laceration groups to be as close as they were while the mean performance of the massage group is, relatively, much different. I thought the chocolate group would do worse than the massage group but the laceration group would do far worse than the chocolate group. Instead, the chocolate group did far worse than the massage group and then the laceration group did slightly worse than the chocolate group. In terms of the age groups, I predicted that mean score would increase with grade level, which it did, and that there might well be a difference in how older subjects cope with their external environment versus how younger subjects cope with their external environment, which there was as well.

6.0 Data Analysis

Table 1 gives an overview of the results before formal statistical analysis. Though ANOVA has not been done yet it appears that there are definitely significant differences in mean scores between both treatment groups and between block groups. The largest are highlighted in red and yellow--0.85 mean for all massage subjects and .51 mean for all laceration subjects and .83 mean for all college grade level subjects and .40 mean for all primary grade level subjects.

| | Primary Block | Elementary Block | Mid. School Block | High School Block | College Block | Treat Means |
|-------------|---------------|------------------|-------------------|-------------------|---------------|---------------|
| Massage | 0.67475 | 0.799 | 0.866625 | 0.932625 | 0.958875 | 0.846375 |
| Laceration | 0.220875 | 0.36425 | 0.52075 | 0.67475 | 0.75 | 0.506125 |
| Chocolate | 0.31125 | 0.435375 | 0.6145 | 0.714375 | 0.77275 | 0.56965 |
| Block Means | 0.4022917 | 0.532875 | 0.6672917 | 0.7739167 | 0.8272083 | Gr mean: 0.64 |

TABLE 1. Mean Results of Each Block and Treatment Group

Table 2 shows the table for the basic ANOVA model, without interaction effects. As predicted, the effects of both the treatments and the different blocks are significant.

| | Deg Free. | Sum Squares | Mean Sum Sq. | F value | Pr(>F) |
|--------------------|-----------|-------------|--------------|---------|------------|
| Between Treatments | 2 | 65.46 | 32.73 | 830.4 | <2e-16 *** |
| Between Blocks | 4 | 73.02 | 18.26 | 463.2 | <2e-16 *** |
| | 2993 | 117.96 | 0.04 | | |

TABLE 2. Effects Analysis

Table 3 shows the ANOVA table for the full fitted model. As depicted, the effects of the treatment factor, the block factor, and their interaction are all significant.

| | Deg. Freedom | Sum Squares | Mean Sum Sq. | F-value | Pr(>F) |
|-----------------|--------------|-------------|--------------|---------|-----------|
| Treatment | 2 | 65.46 | 32.73 | 865.99 | <2e-16*** |
| Block | 4 | 73.02 | 18.26 | 483.01 | <2e-16*** |
| Treatment*Block | 8 | 5.14 | 0.64 | 17.01 | <2e-16*** |
| Residual | 2985 | 112.82 | 0.04 | | |
| Total | 2999 | 256.44 | | | |

TABLE 3. Anova For Full Fitted Model

Table 4 shows the results of Tukey's test on each of the treatments. As shown, all of the groups have statistically significant differences, though the magnitude of the difference between the laceration and chocolate groups is far less than the difference between the massage group and any other group. Tukey's test identifies the difference as the mean score of the chocolate group being 06.3525% greater than the mean of the laceration group. The test is out of a total of forty points, so this translates to 2.541 exam points. By comparison, Tukey's test finds the mean of the massage group to be 11.069 points greater than that of the chocolate group.

| Treatments | Difference | Lower Bound | Upper Bound | P-adj |
|--------------|------------|-------------|-------------|-------|
| Lacer - Choc | -0.063525 | -0.08434371 | -0.04270629 | 0 |
| Mass - Choc | 0.276725 | 0.25590629 | 0.29754371 | 0 |
| Mass - Lacer | 0.340250 | 0.31943129 | 0.36106871 | 0 |

Figure 2 plots these results.

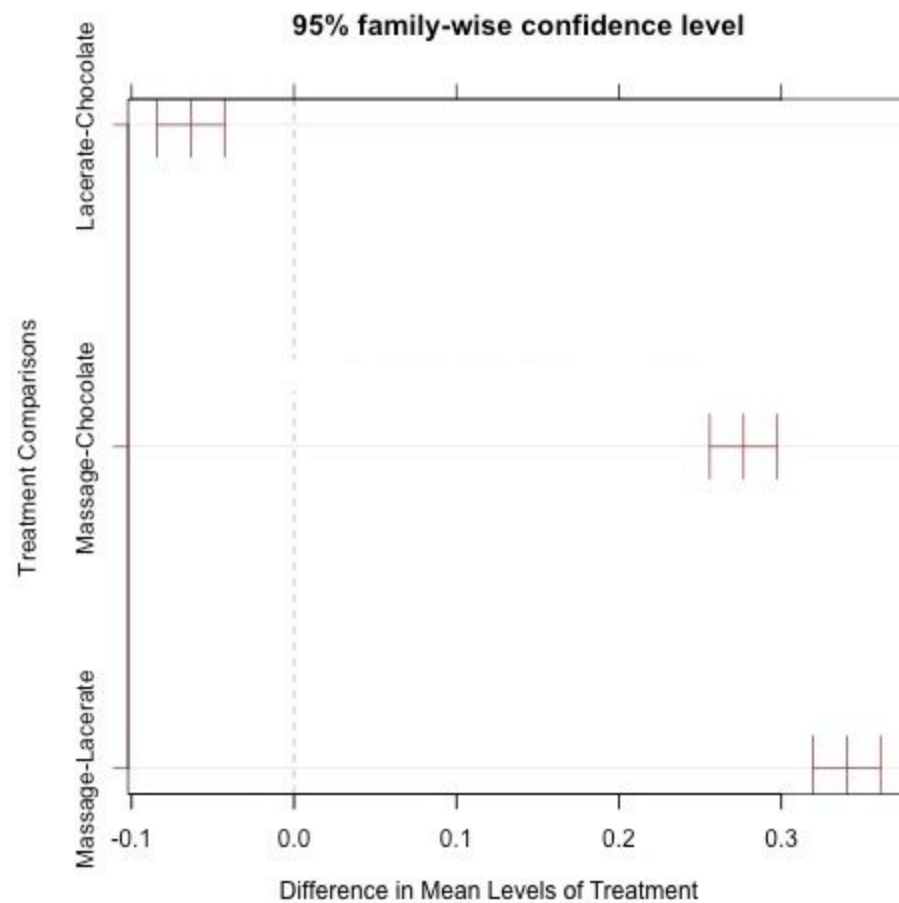


FIGURE 2. Plot of Tukey Test Intervals

7.0 Conclusions and Recommendations

The intent of this experiment was to determine if any differences in academic performance existed between different pre-test environments and experiences. The million question is actually whether or not there are differences in performance between emotional and mental states but it's not possible to directly apply specific emotions as treatment so instead the experiment applied external treatments that then cause the selected emotions. So, we are actually only testing in terms of pre-test environment and experience but, if not applicable in its own right (fortunately lacerated students will probably never be a subject of debate or consideration) this allows us to at least speculate on the effects of emotional and mental well-being on academic performance (laceration reduces self-confidence and if lacerated students perform worse then perhaps increasing students' self-confidence could be worth investing in). Whatever the actual ramifications of the experiment, the design structure, execution, and conclusions remain valid.

Basic intuition suggests that students would perform better after receiving a massage than after an unhealthy snack or after receiving a laceration--essentially torture and a treatment which we will probably find ourselves in court for pretty soon. Indeed, this intuition is consistent with the outcome of the experiment. Notably, however, the difference between massage and chocolate is about 11 points (out of 40) on average whereas the difference between chocolate and laceration is just 2.5 points (out of 40) on average.