

The Effect of Color on Short-Term Memory Recall

Emily Zou¹ and Alison Youndt[#]

¹Conestoga High School, USA

[#]Advisor

ABSTRACT

This study investigates the effect of color on short-term memory retention through a simple flashcard-based assessment. The experiment conducted in Conestoga High School's Unity Day, randomly assigned participants to two treatments: colored or non-colored flashcards. Participants were instructed to recall the ten words presented to them twice within a 30-second interval. A one-sided two-sample t-test and two one-way ANOVA tests were conducted to determine if there was a difference in the memory retention for the colored flashcard and non-colored flashcard groups and if there was a difference in the memory retention amongst 9th, 10th and 11th graders for the colored flashcard and non-colored flashcard groups respectively. The results were statistically insignificant but limited by convenience sampling and premature stops in the experiment. Future points of interest include larger sample sizes, teacher-student differences and the association of academic intellect with memory retention.

Introduction

When I was in sixth grade, my science teacher had challenged our class to remember ten unrelated words in a minute. When he called time, most students recalled only half the words. However, before instructing us to remember these words again and undergo the same test, he told us a story with these words put into context. The next time the students were asked to recite these ten words, most of them, including me, were able to shout out most, if not all of them.

Personally, I was curious about how the brain would store two sources of information at once— the color of the word and the word itself. For example, would a subject be overwhelmed with recalling the color of the word and forget to recall the word itself at the end of the experiment or would they first recall the color, and thus the word associated with it, forming a neurological shortcut in their memory path?

However, there also exists scientifically proven studies that support the association of words with external stimuli, such as color, which will serve as the primary factor of investigation in this study. Color is often regarded as one of the most important visual experiences to human beings and has the potential to enhance memory recall (Adams FM, Osgood CE.). In 1976, researchers Farley and Grant concluded that colored multimedia presentations yielded better attention and memory-performance when compared to their non-colored counterparts (Farley, F.H., Grant, A.P.).

This experiment on memory recall intended to investigate whether the association of color with words, which serves as a potential connection point for the brain, heightens the ability of a subject to recall it after a short period of time. Would a subject be overwhelmed with recalling the color of the word and forget to recall the word itself at the end of the experiment or would they first recall the color, and thus the word associated with it, forming a neurological shortcut in their memory path?

Hypotheses:

1. Subjects who are assigned to the flashcards with colored words will correctly recall more words, on average, than subjects who are assigned to flashcards with non-colored words.

2. There is a difference in the mean number of correctly recalled words between 9th, 10th and 11th grade students who were assigned to the flashcards with colored words.
3. There is a difference in the mean number of correctly recalled words between 9th, 10th and 11th grade students who were assigned to the flashcards with non-colored words.

Methods

The study randomly assigned subjects to one of two treatments, flashcards with colored words and flashcards with non-colored words, and was designed to influence the number of words correctly recalled from the flashcards from subjects.

The experiment was conducted on Unity Day (a tradition in Conestoga High School in which there are no formal classes for the entire day) in the Atrium, an open two-floored gathering space, from periods 2-4, and period 7. The researchers each took on two different shifts each and at least two researchers were present during each period of study. One researcher would read the script and verify the accuracy of responses while another would time the responses of the student.

Since Unity Day occurs after 12th grade students leave, the experiment targeted the population of all 9th grade, 10th grade, and 11th grade students and faculty in the building. A convenience sample of 61 subjects who frequented the table at which the experiment was taken. Thus, participation was completely voluntary and there was no attempt to randomly sample students.

Each subject was instructed to flip a coin to randomize their treatment assignment and allow the determination of the effect of color on short-term memory retention. If the subjects flipped heads, they would be assigned to the non-colored flashcards and if they flipped tails, they would be assigned to the colored flashcards group. A total of 29 responses were collected for the non-colored flashcards treatment and 32 responses were collected for the colored flashcards treatment. To begin the procedure, the researchers would present subjects with ten flashcards that each had a single, distinct word on it. The ten words were all the same for the two treatments to control for the complexity of the words and all words were hand-written by one of the researchers (Mindy Wang) to control for the eligibility of the words. However, for the colored flashcards treatment, each word was written in a different color and the one the researchers most heavily associated with it. For example, the word “hostile” was written in red, which traditionally represents anger and aggression in widespread cultures. On the other hand, for the non-colored flashcards treatment, each word was always written in black. The rest of the procedure was the same for both treatments as described in the script.

In order to not influence other subjects’ responses, a folder hiding a key of the ten words was set up and could be freely referred to by the researchers. In addition, it was advised for those waiting in line behind the active participant to stand around five feet away from them and not look at the flashcards prematurely, which would likely affect their memory retention when they underwent the experiment in the near future.

Materials:

- Pens
- Printed half-sheets for collecting information
- A coin
- Two folders glued together
- A key of ten words
- A phone for timing

Hand-out:

Color and Memory Recall Project

I agree to not tell anyone else what was written on the flashcards or how I participated in the study until after the school day.

Student Teacher

If Student: Grade _____

Circle One: Heads Tails

Words Recalled (Researcher Only):

- | | | | |
|---------------------------------|----------------------------------|---------------------------------|---------------------------------|
| <input type="checkbox"/> Word 1 | <input type="checkbox"/> Word 2 | <input type="checkbox"/> Word 3 | <input type="checkbox"/> Word 4 |
| <input type="checkbox"/> Word 5 | <input type="checkbox"/> Word 6 | <input type="checkbox"/> Word 7 | <input type="checkbox"/> Word 8 |
| <input type="checkbox"/> Word 9 | <input type="checkbox"/> Word 10 | | |

Total Score ____/10

Thank You for Participating!

Script: Thank you for volunteering to participate in our statistical experiment on memory recall. We will show you a series of 10 flashcards, each with a distinct, single word on it and will go through all ten flashcards twice during a period of 30 seconds. Next, we'll wait 15 seconds. Then we will ask you to list all the words you recall, which do not have to be in order, within a 30 second interval. Would you like us to repeat these instructions?

Before we begin, please fill out this form and check the box at the bottom to agree that you will not tell any other students what was written on these flashcards after the statistical experiment is over.

Great, let's begin. First, flip this coin.

We will start flipping through the flashcards now.

(30 seconds pass)

Time is up. Please wait another 15 seconds before answering.

(15 seconds pass)

Time is up. Tell us all the words you remember. You do not have to recite them in order.

(30 seconds pass)

Thank you for participating in our experiment! Remember to not tell anyone else what was on these cards until the school day is over."

Key:

First Word: Investigation – Grey

Second Word: Abbey – Pink

Third Word: Kettle – Black

Fourth Word: Community – Green

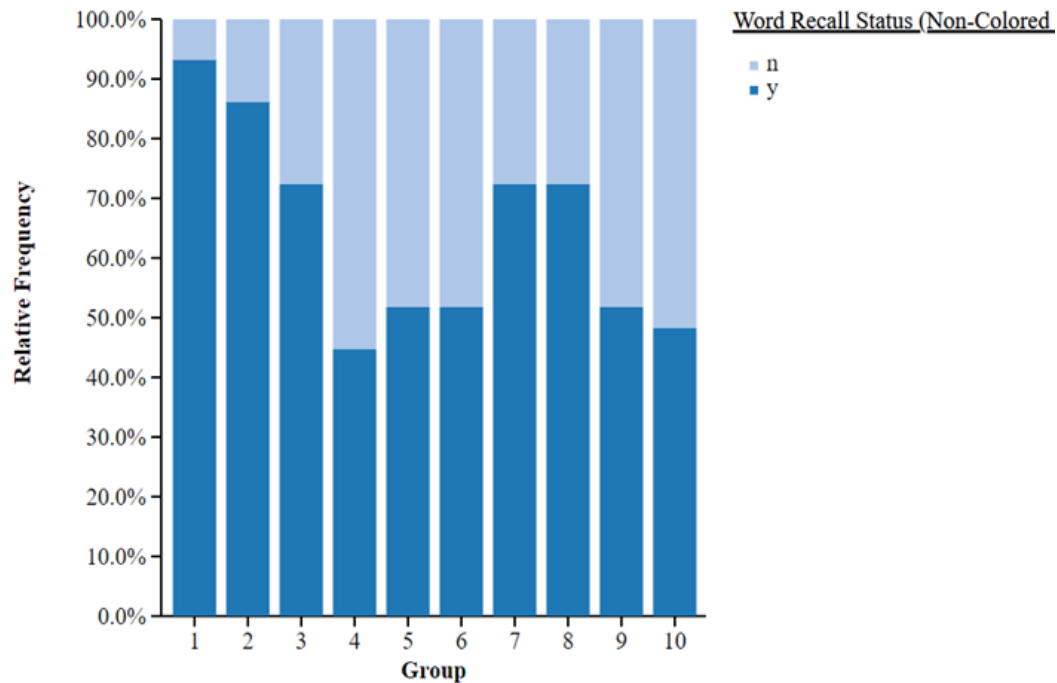
Fifth Word: Psychology – Yellow

Sixth Word: Captivate – Purple
 Seventh Word: President – Brown
 Eighth Word: Soap – Blue
 Ninth Word: Hostile – Red
 Tenth Word: Wing – Orange

Analysis and Results

Distribution of correct answers for each word:

Note: Words are labeled with the number they correspond to in the answer key (*see above). Perhaps unsurprisingly, the word most often correctly recalled was the first word to be shown, or investigation, with almost 95% of participants correctly recalling it. Trailing behind investigation as the second most commonly recalled word was abbey, the second word to be shown, with about 85% of participants correctly recalling it. The most missed word was community, which was shown in the middle of the experiment, and less than half of participants correctly recalled it. Participants often struggled to recall words that were shown at the *middle* and end of the experiment the most.



μ_C : the true mean number of correctly recalled words using the flashcards with colored words.

μ_{NC} : the true mean number of correctly recalled words using the flashcards with non-colored or black words.

$$H_0: \mu_C = \mu_{NC}$$

$$H_a: \mu_C > \mu_{NC}$$

The conditions for the random assignment of treatments, normality of the sampling distributions of \bar{x}_{9C} , \bar{x}_{10C} , and \bar{x}_{11C} , and the independent of subject responses were satisfied.

Table 1. One-sided two-sample t-test for $\mu_C - \mu_{NC}$ using $\alpha = 0.05$:

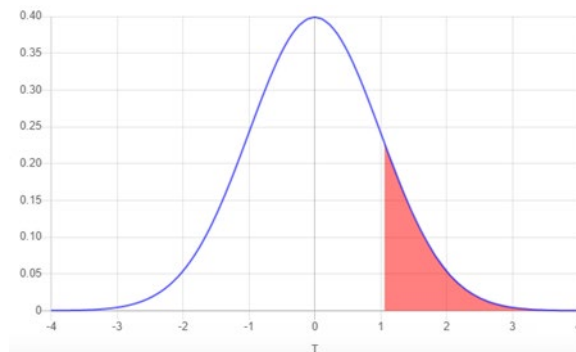
Variable	Observations (n)	Sample Mean:	Std. error
Colored	32	6.875	1.408
Non-Colored	29	6.448	1.703

t-value = 1.06

Degrees of Freedom: 54.549

$H_a = \mu_C > \mu_{NC}$

P-value = 0.147



μ_{9NC} : the true mean number of correctly recalled words by 9th graders using the flashcards with non-colored words.

μ_{10NC} : the true mean number of correctly recalled words by 10th graders using the flashcards with non-colored words.

μ_{11NC} : the true mean number of correctly recalled words by 11th graders using the flashcards with non-colored words.

$$H_0: \mu_{9NC} = \mu_{10NC} = \mu_{11NC}$$

H_a : At least one of the population means are different.

Independence of subject responses and the sampling distributions for \bar{x}_{9NC} , \bar{x}_{10NC} , and \bar{x}_{11NC} were approximately normal. However, no random sampling of students was conducted, and the quotient of the largest standard deviation and the was larger than 2. Thus, the procedure should proceed with caution!

Table 2. One-way ANOVA test of variance using $\alpha = 0.05$:

Variable	Observations (n)	Sample Mean	Std. Error
9 th Graders (Non-colored)	8	6	1.69
10 th Graders (Non-colored)	10	6.5	1.78
11 th Graders (Non-colored)	7	6.143	0.69

Mean Squares for Groups (MSG): 0.601 Degrees of Freedom (denominator): 2

Mean Squared Error (MSE): 2.334 Degrees of Freedom (numerator): 22

F-value = 0.258 P-value = 0.775

μ_{9C} : the true mean number of correctly recalled words by 9th graders using the flashcards with colored words.

μ_{10C} : the true mean number of correctly recalled words by 10th graders using the flashcards with colored words.

μ_{11C} : the true mean number of correctly recalled words by 11th graders using the flashcards with colored words.

$$H_0: \mu_{9C} = \mu_{10C} = \mu_{11C}$$

H_a : At least one of the population means are different.

The conditions for independence and standard deviations were satisfied. However, the convenience sampling of subjects did not satisfy the random sampling condition and the sampling distributions of the sample means of \bar{x}_{9C} , \bar{x}_{10C} , and \bar{x}_{11C} were skewed.

Table 3. One-way ANOVA test of variance using $\alpha = 0.05$:

Variable	Observations (n)	Sample Mean	Std. Error
9 th Graders (Colored)	6	6.667	1.211
10 th Graders (Colored)	7	6.857	0.9
11 th Graders (Colored)	16	7.125	0.957

Mean Squares for Groups (MSG): 0.513 Degrees of Freedom (denominator): 2

Mean Squared Error (MSE): 0.998 Degrees of Freedom (numerator): 26

F-value = 0.514 P-value = 0.604

Conclusion and Discussion

None of the inference procedures yielded statistically significant results and it couldn't be concluded that the association of colors with words or the age of a student affected a subject's ability to recall words after a short period of time.

For the two-sample one-sided t-test, since the p-value = .147 > $\alpha = .05$, we fail to reject the null hypothesis and don't have convincing evidence that true mean number of correctly recalled words is greater for the group assigned to the flashcards with colored words than the group assigned to the flashcards with non-colored words for subjects like those in the study. These results could only be generalized to subjects like those in the study rather than all 9th, 10th and 11th grade students and faculty in Conestoga High School. For the one-way ANOVA Test of variance for the non-colored flashcards group, since the p-value = .775 > $\alpha = .05$, we fail to reject the null hypothesis and don't have convincing evidence that there is a difference in the true mean number of correctly recalled words between the 9th, 10th and 11th graders like those in the study assigned to the non-colored flashcards treatment. Likewise, for the one-way ANOVA Test of variance for the colored flashcards group, since the p-value = .604 > $\alpha = .05$, we fail to reject the null hypothesis and don't have convincing evidence that there is a difference in the true mean number of correctly recalled words between the 9th, 10th and 11th graders like those in the study assigned to the colored flashcards treatment. Both ANOVA Tests could only be generalized to 9th, 10th and 11th grade subjects like those in the study.

Limitations

The experiment relied on several assumptions, which led to inevitable shortcomings. Most notably, a convenience sample was taken from the population of all students and faculty in Conestoga High School, which prevented the generalizations of the findings to the target population. Convenience sampling also played the most disruptive role in the ANOVA test of variance for 9th, 10th and 11th graders that were assigned to the colored flashcards group and in the ANOVA test of variance for 9th, 10th and 11th graders that were assigned to the non-colored flashcards group. While the treatment of the flashcards could be randomized, the sampling of students from each grade could not be randomized and thus, could not fulfill the random sampling criteria required by the ANOVA inference procedure. Consequentially, both ANOVA tests were advised to proceed with caution. Because of the convenience sample, there existed unequal sample sizes for each grade; for example, in the ANOVA test for the colored flashcards group, there were far more 11th grade students than 9th or 10th grade students, likely since the majority of AP Statistics students were juniors themselves and often frequented the table when they were done with their own shifts. For smaller samples, there existed more variation within the groups themselves rather than amongst the groups, producing smaller F-values that yielded insignificant results. In the future, standardization of the sampling procedure to include random sampling of subjects based on grade and larger sample sizes would be suggested.

Another shortcoming was preventing students who stood in line behind the active participant from seeing the flashcards or hearing the words recited prematurely. Since the participant was not allowed to view the researchers' side of the folder hiding the script and key, participants were forced to undergo the experiment with their backs turned against the line, which would allow the line to quickly glance at the presented flashcards. Though it was advised for potential subjects to stay back at least three feet from the table, it was hard to maintain that rule as new people filed in. However, those who did see or hear what was on the flashcards and then participated would have an inflated number of words they correctly recalled, regardless of which treatment they were assigned to. In the future, it would be suggested to uphold the principles of a single-blind experiment and limit extra exposure to the words by setting strict guidelines on where to stand in line or ideally, call subjects in one by one to a different room where the experiment was done.

An easily preventable shortcoming was the premature ending of. Although a 30-second period was allocated for subjects to recall all the words they remembered from the flashcards, not every subject used the full 30 seconds, and some withdrew their participation from the experiment prematurely. Out of the interest of time and the acknowledgement that these subjects were likely not going to exert more brainpower in recalling words they didn't remember; the experiment was ended accordingly. However, there was one trial at the end of Unity Fair in which the subject was forced to use the full 30 seconds for recitation and recalled one more word after supposedly giving up. Thus, by allowing subjects to end the experiment early, the researchers unintentionally influenced the subjects' responses and understated the number of words they could recall in thirty seconds. In the future, the script should be more strictly followed, and the full 30-second recitation period should be allotted, even if it is against the participant's wishes.

On the other hand, some of the study's strengths were more active engagement from the participants—flipping a coin to randomize their own treatments, filling out a form, challenging their memory by asking to recall as many words as possible— and a relatively simple randomization process. The design of the experiment was also simple and easy to execute since there were only two levels associated with the explanatory variable. Participation was also high since the experiment was easily understandable and on the shorter side.

Future points of interest to investigate are if there is a difference in the true mean number of words recalled between students and teachers at Conestoga High School and if there is a linear relationship between the amount of AP classes a student enrolls in (as an approximate measure of academic intellect) and the number of correctly recalled words.

Acknowledgments

I would like to thank my fellow researchers Mindy Wang and Valerie Xu for helping to carry out the experiment in the most objective way possible and my advisor Allison Youndt for her academic support and encouragement while writing this paper.

References

- Adams FM, Osgood CE. A cross-cultural study of the affective meaning of color. *J Cross Cult Psychol.* 1973;4(2):135–156.
- Farley, F. H., & Grant, A. P. (1976). Arousal and cognition: Memory for color versus black and white multimedia presentation. *The Journal of Psychology: Interdisciplinary and Applied*, 94(1), 147–150. <https://doi.org/10.1080/00223980.1976.9921410>