

History Knowledge Presented by Different Media Predict the Learning Efficacy of Students in a Short Period

Jiaming Xu¹, Junzhi Wei², Muzi Li³, Yang Xia⁴, Hongyu Hu⁵

¹Shanghai Southwest Weiyu Middle School, Shanghai, 200233, China.

²Stuart Hall School, Virginia, 24401, United States.

³Catholic Memorial High School, Waukesha, 53186, United States.

⁴The Pennington School, New Jersey, 08534, United States.

⁵Severn School, Maryland, 21146, United States.

Abstract

An online double-blind, randomized controlled experiment in the subject of the learning efficacy of AP history, a system of providing high school students opportunities to participate in advanced courses that are usually offered in universities, was conducted. The experiment was designed to find the most effective means for students to study history contents in online learning. Some students tend to use videos on websites such as YouTube to learn instead of reading the textbooks. The intervention limited the means of studying students to either reading only, watching videos without subtitles, or watching videos with subtitles. Results reveal an insignificant relationship between the media presenting the contents and the learning efficacy, meanwhile significant relationships between the initial interest in history and the learning efficacy and between the prior knowledge in the history and the learning efficacy.

Keywords

Media, learning efficacy, online learning, history learning.

1. Introduction

On video platforms, especially YouTube, numerous people published video lectures about multiple classes; for instance, several channels posted content about AP US History, which is one of the most popular AP classes taken by high school students throughout the US. With the utilization of the Internet becoming more and more common, especially during the outbreak of COVID-19, the means of learning is becoming less and less limited to the traditional ways of learning in the classroom. The most popular means of self-learning are reading textbooks and watching videos online. In the experiment, participants were given either texts or the video form of the same contents with PowerPoint slides to learn. The participants who watched videos were divided into two groups: the group attending the video with subtitles, and the group watching the video without subtitles. The contents of the texts were selected from one of the AP European History textbooks, consisting of three passages.

1.1. AP and IB

AP stands for Advanced Placement, a system created by the educational organization College Board for high school students to learn the contents of the classes that are usually offered in universities. AP exams take place in May every year, and the scores range from 1 to 5, with 5 being the highest score. A high score in the AP exam may help high school students to earn credits in universities early. IB stands for International Baccalaureate and is similar to the AP system. Finishing high-level IB courses could count as college credits in universities.

1.2. European History

AP European History is not a widely offered class in high schools, compared to AP classes such as AP US History, AP US Government and Politics, and AP Calculus. Even in the schools that offer advanced European history classes, it is not as popular as other AP history classes. Meanwhile, the time available for the experiment from being proposed, planned, conducted, and analyzed is within five weeks, which limits the ability to find enough participants for the research. Therefore, considering that most of the potential participants are Chinese high school students studying in the United States, AP European History should be a class most of them were unfamiliar with, with reasons listed above. Therefore AP European History was selected as the topic to research and to ask the participants to learn.

1.3. Hypotheses

Prior studies have shown that watching videos improves the performance of the participants in the test after learning, and better predicts the performance as well [1]. Therefore, the first hypothesis took the prior research results into account and predicted that the video form of historical knowledge would help the participants to achieve higher learning efficacy.

H1: Students who watch videos will score higher in the post-test than those who read texts only in a short period.

EFL (English as a Foreign Language) undergraduates, which in this case, is similar to the potential participant population, understood the English narrations in the videos more thoroughly with the assistance of English subtitles [2, 3]. Therefore, the second hypothesis predicts that the English subtitles in the videos will improve the learning efficacy of the students.

H2: Students who watch videos with subtitles will score higher in the post-test than those who watch videos without subtitles in a short period.

Krapp [4] found that the initial interest of the participants in the topic of the learning contents influences their attitudes towards different means to learn and their specific use of learning strategies in concrete learning situations. Schiefele [5] also suggested that there is a strong correlation between initial interest and achievement. The initial interest of the participants in history was taken into account in the study, as well as two other variables: the prior knowledge of the participants in the contents reflected by the pretest scores, and the gender of the participants.

Sub-hypothesis 1: Students who have a higher initial interest in history will score higher in the post-test than those with little or no initial interest in a short period.

Sub-hypothesis 2: Students who have higher scores in the pretest will score higher in post-test.

Sub-hypothesis 3: Female students will score higher reading texts, while male students will score higher watching videos.

2. Method

2.1. Participants

The total number of participants who took part in the study is 65 people. The gender ratio of participants is balanced that we have 33 males and 31 females (1: 0.94). The samples participants include all educational level students: freshman(n=13), sophomore(n=14), junior (n=15), senior (n=20), and above 12 grades (n=3). The most basic criteria that all these participants meet include that they have a certain level of English learning and intend to go abroad or have already studied abroad and haven't taken AP European History.

The Sampling method we used is the Random Sampling Method and Random Assignment Method. We employed a 'Research Randomizer' website to randomize our sample. First, the

sample participants will receive an initial code from 1 to 65, and then all these numbers are put into this randomizer resulting in three different groups. In the following experiment, there will be a text group, video without subtitles, and video with subtitles group. Later, participants will have a new code in each group based on their sequence in the randomized group on the website. Besides, each group will have a unique prefix in their system. The text group has code starting with TE. Numbers in the video without subtitles group have code starting with AN, and video with subtitles group starts with VD. As a result, each experimental group has an equal number of participants. TE and AN group have 22 participants, and the VD group has 21 participants. Participants are randomly assigned to the group of either text media or video media. No participants will have both the experience of these two media.

2.2. Procedure

Due to the pandemic, the whole experiment was done online with the participants. There were two phases. The first phase was inviting our participants to join the study. Participants for the study were recruited from our high school friends. Participants were introduced to our research, general procedure, the duration, and the purpose of this study, not the real one but related one, which was whether our adaptation of the history textbook is appropriate and helpful for students to learn. The reason that encouraged participants to join the study was that they could learn a lot about a new subject that may help them with future performance in school. Finally, they received WeChat Money at the end of the experiment that expresses our appreciation.

The second phase was the experimental stage. Participants received a link of the consent form as a sign of agreement to participate in the experiment. After that, all the participants are randomly assigned into three groups, and each participant received an identity code that would be used in later surveys. Then participants received a background questionnaire and pretest. The Background questionnaire contained questions about gender, year of birth, subject, and level of English. The real intention was to know the participants' initial interest in history, which could be used to analyze hypotheses. The pre-test tested participants some knowledge in European history, which showed participants' familiarity with European history that helped to be used for hypothesis analysis. Participants then read a text, watched a 5 min video without subtitles, or watched a 5 min video with subtitles to learn some European history knowledge. No matter which way of three they learned from, there was a 7-minute time limit that controlled the time of learning for every participant the same, and participants could review the content if there were a remaining time. Participants needed to be connected to the experimenter through Wechat call. They were not allowed to mute the sound so that it may avoid cheating and solve some technical issues immediately. When learning the knowledge, participants could not take notes which controlled the testing environment, and supplies are the same for everyone. Finally, participants would be informed of a test after learning to pay more attention when learning. After learning, participants received a link to the post-test. There was a 7-minute time limit, but they could submit early if they wanted. Questions in the post-test correspond to the content participants have learned in the previous stage. Participants received a debriefing form. It told the participants the real purpose and study we did and why we did not tell them the real intentions before. At the end of the experiment, each participant received Wechat money to express our appreciation.

2.3. Measurements

The independent variables which are manipulated in the experiment include the way that participants learn: text media and video media. The dependent variable is the participants' performance on the post-test. The other naturalistic measurements in the experiment are participants' English learning level, their taken or favorite subjects, their favorite movies, and their gender.

A short period of time

The experiment will last about 15 minutes, including 7 minutes to read or watch, and another 7 minutes to do the post-test, which is a short time for students to learn our content.

Learning efficacy

It will be reflected by how many scores that students get in the post-test. The results of this measurement can be directly exported from the Qualtrics.

Initial interest in history

Whether the participant selects history as his or her favorite subject(s), movie(s), or AP courses except for AP European History, the more he or she selects, the more he or she is likely to be interested in history. So if the participants choose any of these options, we assume they have Initial interest in history.

Text media

Only text or online written forms are involved in the teaching materials. Participants will receive a text document (about two pages) that contains all three passages. Each passage contains around 300 words, and all of our content is from the book, *Western Civilization 10th ed* [6].

Video media

There are sounding recordings, pictures, slides, and animations in the teaching materials. Participants will receive a YouTube or Bilibili private link of a video that contains all the contents in the text. There are two types of videos: one contains subtitles; the other one doesn't. Both videos have the same duration, 5 minutes, and 10 seconds.

2.4. Preliminary Analysis

The first step of analyzing data was filtering erroneous data. Based on the background survey, there were three samples unqualified. Sample with code AN006 was not qualified because the participant had taken AP European history. The reason that none of the sample participants should take this course before was that the content they would learn in the experiment was relevant to European history. To limit the bias of results, this sample participant would not be analyzed. Sample with code VD005 was also excluded because the participant preferred not to define gender. Since gender was a measurement related to one of the hypotheses, this sample could also not be analyzed. The participant with code VD017 was not qualified because of his English learning level. English learning level was one of the critical naturalistic measurements because all the contents, including text, video, and test, were in English. Thus, participants were assumed to have a qualified English learning level, shown by TOEFL scores, to understand contents. The measurements described in the measure part included the English learning level, the courses were taken, and favorite classes, the preferred types of movies, and the gender of the participants to ensure that all participants met the basic criteria.

2.5. Statistical Analysis

For the main hypotheses, the mean score of the post-test of each group is calculated (Table 1). The comparison between different groups, including the text group (shown as TE), video without subtitles group (shown as AN), and video with subtitles (shown as VD), is clearly shown in Figure 1.

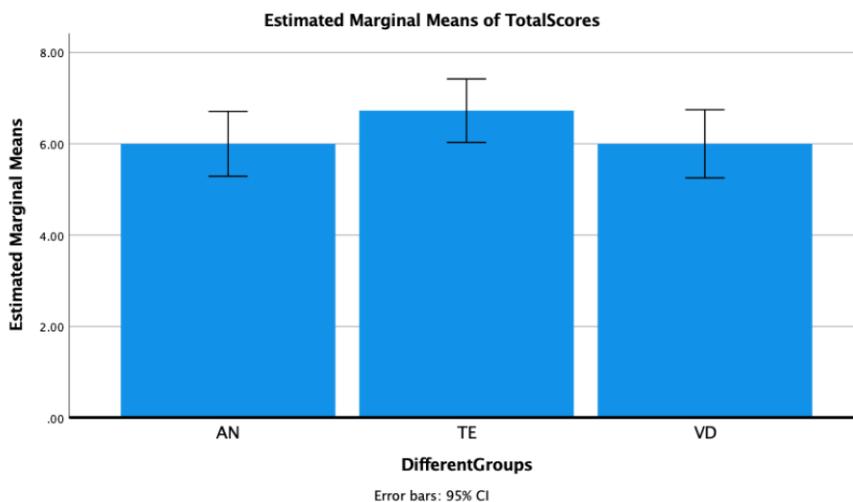


Figure 1. Mean scores of the post-test of participants in three manipulated groups. Error bars represent 95% confidence intervals

Table 1. Mean scores, standard deviation and number of three manipulated groups in post-test

DifferentGroups	Mean	Std. Deviation	N
AN	6.0000	1.18322	21
TE	6.7273	1.80428	22
VD	6.0000	1.82574	19
Total	6.2581	1.63903	62

Table 2. Tests of between subjects effect

a. R Squared = .046 (Adjusted R Squared = .013)

b. Computed using alpha = .05

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power ^b
Corrected Model	7.507a	2	3.754	1.416	.251	.046	2.833	.292
Intercept	2406.989	1	2406.989	908.219	.000	.939	908.219	1.000
Different Groups	7.507	2	3.754	1.416	.251	.046	2.833	.292
Error	156.364	59	2.650					
Total	2592.000	62						
Corrected Total	163.871	61						

For sub-hypothesis 1, the mean scores of the post-test of participants who are interested or not interested in history are calculated (Table 3). The comparison between different groups is clearly shown in Figure 2.

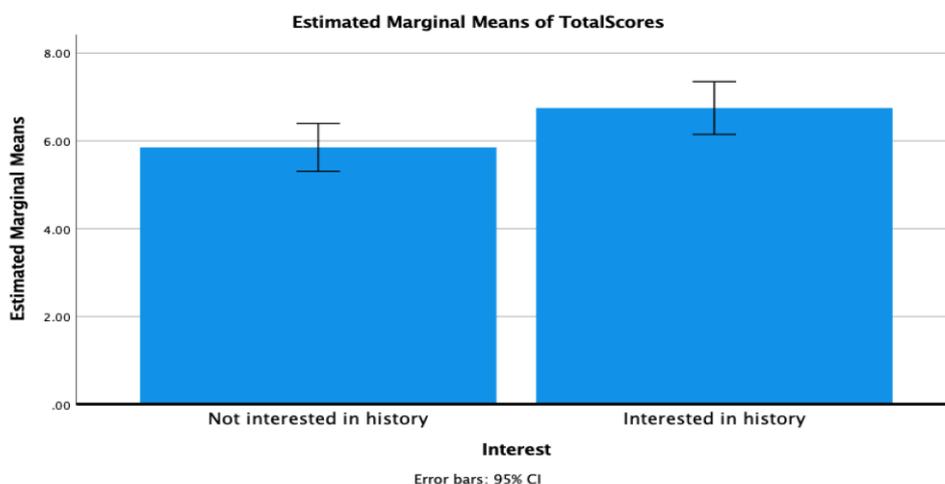


Figure 2. Mean scores of the post-test of participants who are interested or not interested in history. Error bars represent 95% confidence intervals.

Table 3. Mean scores, standard deviation and the number of post-test of participants who are interested or not interested in history.

Interest	Mean	Std. Deviation	N
Not interested in history	5.8529	1.65387	34
Interested in history	6.7500	1.50616	28
Total	6.2581	1.63903	62

Table 4. Tests of between subjects effect

a. R Squared = .075 (Adjusted R Squared = .060)

b. Computed using alpha = .05

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power ^b
Corrected Model	12.356a	1	12.356	4.893	.031	.075	4.893	.586
Intercept	2438.872	1	2438.872	965.796	.000	.942	965.796	1.000
Different Groups	12.356	1	12.356	4.893	.031	.075	4.893	.586
Error	151.515	60	2.525					
Total	2592.000	62						
Corrected Total	163.871	61						

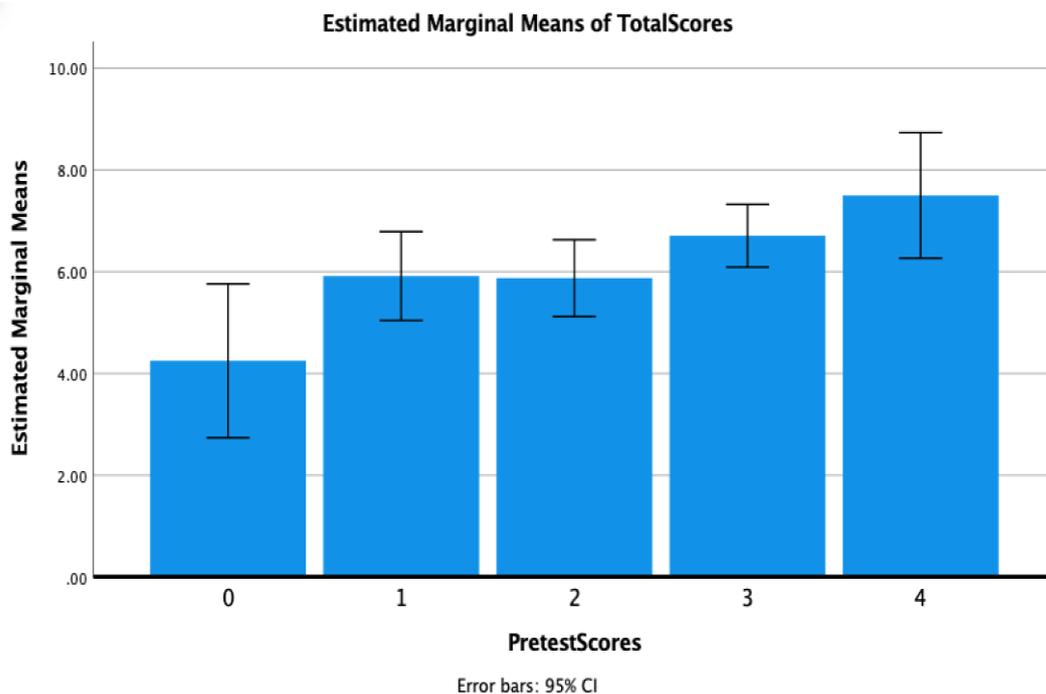


Figure 3. Mean scores of the post-test of participants with different pre-test scores. Error bars represent 95% confidence intervals

For sub-hypothesis 2, the post-test mean scores of participants with different pre-test scores are all shown in Table 5, which will provide precise numbers, while the increasing trend of post-test mean score is exhibited in Figure 3.

Table 5. Mean scores and standard deviation of participants with different pre-test scores

PretestScores	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0	4.250	.755	2.739	5.761
1	5.917	.436	5.044	6.789
2	5.875	.377	5.119	6.631
3	6.708	.308	6.091	7.325
4	7.500	.616	6.266	8.734

For sub-hypothesis 3, there are two variables (gender and learning media). The mean scores of different genders in each group are presented in Table 7. Besides, the comparison between six different conditions can be easily found in Figure 4, which presents not only the difference between participants in the same gender but with different media, but also that between participants with the same medium but in different genders.

Table 6. Tests of between subjects effect

a. R Squared = .207 (Adjusted R Squared = .152)

b. Computed using alpha = .05

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Powerb
Corrected Model	33.996a	4	8.499	3.730	.009	.207	14.920	.859
Intercept	1514.586	1	1514.586	664.727	.009	.921	664.727	1.000
Different Groups	33.996	4	8.499	3.730	.009	.207	14.920	.859
Error	129.875	57	2.525					
Total	2592.000	62						
Corrected Total	163.871	61						

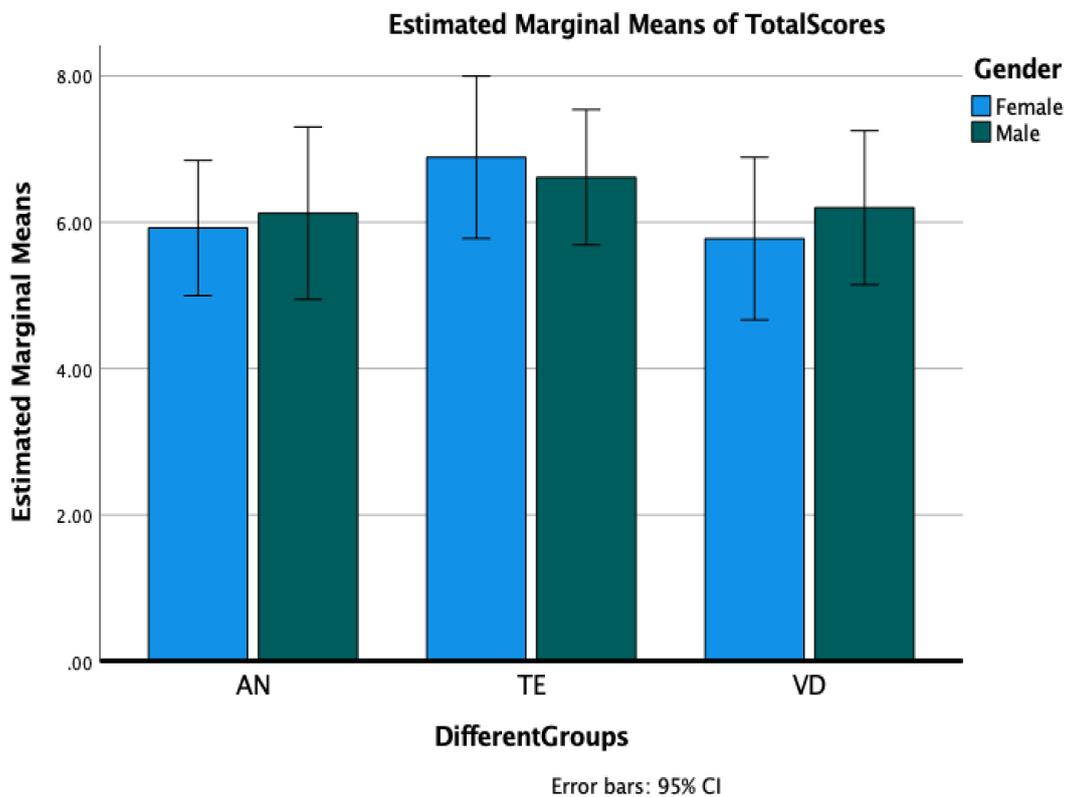


Figure 4. Mean scores of the post-test of participants of different genders and three manipulated groups. Error bars represent 95% confidence intervals

Table 7. Mean scores and standard deviation of participants with different pre-test scores

DifferentGroups	Gender	Mean	Std. Deviation	N
AN	Female	5.9231	1.18754	13
	Male	6.1250	1.24642	8
	Total	6.0000	1.18322	21
TE	Female	6.8889	1.69148	9
	Male	6.6154	1.93815	13
	Total	6.7273	1.80428	22
VD	Female	5.7778	2.22361	9
	Male	6.2000	1.47573	10
	Total	6.0000	1.82574	19
Total	Female	6.1613	1.69503	31
	Male	6.3548	1.60309	31
	Total	6.2581	1.63903	62

Table 8. Tests of between subjects effect

a. R Squared = .055 (Adjusted R Squared = -.030)

b. Computed using alpha = .05

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power ^b
Corrected Model	8.952a	5	1.790	.647	.665	.055	3.236	.218
Intercept	2343.345	1	2343.345	847.068	.000	.938	847.068	1.000
Different Groups	7.641	2	3.820	1.381	.260	.047	2.762	.285
Gender	.205	1	.205	.074	.787	.001	.074	.058
DifferentGroups * Gender	1.285	2	.643	.232	.794	.008	.465	.085
Error	154.919	56	2.766					
Total	2592.000	62						
Corrected Total	163.871	61						

3. Results

For the main hypotheses, the mean score of the group with text media is the highest one, which is 6.7273 points (Table 1). The mean scores of the two video groups are the same, which is 6.0000 (Table 1). The statistical results of the experiment are contradictory to our expectation, which H1 predicts that students who watch videos will score higher in the post-test than those who read texts only in a short period, and H2 predicts that students who watch videos with subtitles will score higher in the post-test than those who watch videos without subtitles in a short period. In addition, the p-value is greater than .05, showing that it is an insignificant result. Therefore, H1 and H2 are both not proved.

For sub-hypotheses 1: Students who have a higher initial interest in history will score higher in the post-test than those who have little or no initial interest in a short period, the result shows that the post-test mean score of those who are interested in history is 6.7500, about 1 point higher than that of those who are not interested in history, which is 5.8529 (Table 3). This is in accordance with our expectations. The p-value is also calculated to be .031 (<.05), which can support that this result is significant. In conclusion, our sub-hypothesis 1 is supported.

For sub-hypothesis 2: Students who have higher scores in the pretest will score higher in post-test, the result (Table 5) shows that those who get 0 points in pretest only get 4.25 points in the post-test; those who get 1 point in pretest get 5.917 points in post-test; those who get 2 points in pretest get 5.875 points in post-test; those who get 3 points in pretest get 6.708 points in post-test; and, those who get 4 points in pretest get 7.500 points in the post-test. It is clear that there is an increasing trend of the mean post-test score when the pretest score is higher, except one counterexample, which is the one of those who get 2 points in the pretest (Table 5). The possible reason for this counterexample is luck, since participants are likely to answer questions correctly by guessing. In general, the increasing trend is consistent with our hypothesis. Also, the p-value is .009(<.05), which strongly supports our hypothesis.

For sub-hypothesis 3: Female students will score higher reading texts, while male students will score higher watching videos, the data itself appears to be consistent with this hypothesis. Regarding reading texts, female participants' mean score is .274 higher than the male participants' mean score, while male participants' mean score is .202 and .422 higher than female participants' one when they watch videos without subtitles and videos with subtitles respectively (Table 7). However, the p-value of this sub-hypothesis is bigger than .05, which indicates that this sub-hypothesis is not significant.

4. Conclusion and Discussion

Conclusion about Hypothesis 1

This study shows that the media does not have a significant effect on learning efficacy. Although our participants reading texts scored higher on average than the ones watching videos did, the correlation turns out to be insignificant. However, even if this correlation eventually becomes significant, it will not be consistent with the findings from previous studies. Choi, H., & Yang, M. found that students who watch problem-based video instructions gain more satisfaction, empathy, and learning achievement than students who read problem-based text instructions[1,3]. Plus, Wiggen, J., & McDonnell, D. shows that students watching videos performed better on their tests and predicted their grades more accurately than students reading texts[7].

Conclusion about Hypothesis 2

This study shows that subtitles in videos do not affect the learning efficacy of the students. In our data, students watching videos with subtitles received the same average score of students watching videos without subtitles, which also eventually turns out insignificant. In contrast, H.

Lin, Yuan-Husan Lee, Dai-Yi Wang, & Sunny S. J. Lin. suggests that EFL (English as a Foreign Language) undergraduates watching videos with subtitles performed better and reported a lower cognitive load when listening to the English narrations [2]. Also, Yang, H. 's study shows that subtitles assisted EFL learners' comprehension when listening to CNN news reports[3].

Conclusion about Sub-Hypothesis 1

According to this study, a student's initial interest does influence their learning efficacy, which is consistent with findings from previous studies. In one study conducted by Krapp [4], they found that interest influences students' attitudes towards different types of learning strategies and their specific use of learning strategies in concrete learning situations. Schiefele [5] also suggested that there is a correlation between interest and achievement, which is higher and more stable than the one between other motivations and achievements.

Conclusion about Sub Hypothesis 2

This study shows that there is a causal relationship between a student's pretest score and his or her post-test score.

Conclusion about Sub Hypothesis 3

According to the study, gender does not influence learning efficacy in reading texts and watching videos.

5. Limitation & Future Direction

There are a few limitations in this study, which might cause the insignificance of both the main hypotheses' results.

First, there is no evaluation on our pretest, post-test, or video. We do not know if the pretest reflects the prior knowledge of the participants and if the post-test reflects their efficacy of learning; We only have four questions on the pretest, and they did not cover the points 12 questions in our post-test are testing, which might cause the pretest and post-test testing the participants on imbalanced content. Besides, some of our post-test problems need tremendous logic reasoning, which is extremely difficult to complete without the texts and videos. We also lack knowledge about if the videos reflect the exact content on the test and if they are made in a quality that does not influence student's learning efficacy. The content of our video is conveyed by a person simply reading the textbook with no further explanations, a feature that deprives the advantage of videos and is scarcely used. Moreover, the difficulties of different sections in the post-test might be different, which will affect the results. According to our data, 35 and 30 percent of participants receive a score of 0 or 1 in passage one and passage two, respectively, while only 8 percent of them receive a 0 or 1 in passage 3. Some participants also directly reported to us that the first two sections of the post-test are too hard.

Moreover, our sample size is too small. We have 65 participants in total and 11 potential control variables like grades, gender, favorite type of films, related AP classes that were taken. Many of them require a larger sample size to show significance. Also, conducting our experiments online has many problems. We cannot control the place and time of our experiments, which will influence the conditions of our participants and thus the results: some of them might just have walked out of a math class while some others might do our experiment in bed. We can neither prevent their potential disruptive behaviors such as cheating and viewing distracting content during the experiment, like looking at other websites. Last but not least, our monetary reward might also be a covariate.

For future research, we suggest the researchers hiring students and professionals to evaluate their pretests, post-tests, and videos. They should have a much larger sample than ours and conduct their experiments face to face at the same time. Plus, they should replace the monetary reward with credits or, if they could, nothing.

References

- [1] Choi, H., & Yang, M. (2011). The effect of problem-based video instruction on student satisfaction, empathy, and learning achievement in the Korean teacher education context. *Higher Education*, 62(5), 551-561.
- [2] H. Lin, Yuan-Husan Lee, Dai-Yi Wang, & Sunny S. J. Lin. (2016). Reading Subtitles and Taking Enotes While Learning Scientific Materials in a Multimedia Environment: Cognitive Load Perspectives on EFL Students. *Journal of Educational Technology & Society*, 19(4), 47-58. *prehension Skills. CALICO Journal*, 31(3), 345-373.
- [3] Yang, H. (2014). The Effects of Advance Organizers and Subtitles on EFL Learners' Listening Comprehension. *Journal of Educational Technology & Society*, 17(4), 47-58.
- [4] Krapp, A. (1999). Interest, motivation and learning: An educational-psychological perspective. *European Journal of Psychology of Education*, 14(1), 23-40.
- [5] Schiefele, U. (1996b). Topic interest, text representation, and quality of experience. *Contemporary Educational Psychology*, 21, 3-18.
- [6] Spielvogel, J. J. (2018). *Western civilization* (10th ed.). Cengage Learning. <https://ebooks.cengage.com>
- [7] Wiggen, J., & McDonnell, D. (2017). Geoscience Videos and Their Role in Supporting Student Learning. *Journal of College Science Teaching*, 46(6), 44-49.