Effects of Test Preparation in Learning

Gabe Rojas-Westall
Northwestern University

Deniz Sagnaklar
Northwestern University

Author Keywords
Eye-tracker; fixation map; Tobii Eye Tracker; Eye Tribe; assessment test; test preparation; reading behavior; prior knowledge; gaze; fixation map

Abstract

In this study, we looked at the differences between test preparation and leisure reading. Our motivation was to analyze the effects of prior knowledge of an assessment test on reading behavior, and the consequent performance of students. Our intention was to understand how test preparation affected the process of learning. Our participants were undergraduate students from Northwestern University who had no prior knowledge about the reading they were given. We split our participants into two groups, one who had prior knowledge of an assessment test at the end of the study, and one who did not. We measured total reading time and collected gaze data using an Tobii Eye Tracker 4C. After the reading, we conducted a small post-test to assess what information the students had absorbed. In our data analysis we found that students who had prior knowledge of a test had almost similar reading times to those who did not. However they scored on average 31% better on the assessment test than those who did not have prior knowledge of th test, and the fixation maps showed that students with prior knowledge spent more time on new terms, definitions and numerical data during the reading.

Introduction

The motivation of our study was to analyze the effects of assessment tests on how undergraduate students study and retain information. This motivation arose while conducting background research on the learning patterns of undergraduates. To learn more about how students study and learn, we interviewed prospective participants for our study. We found that many of them were concerned with whether there was a test at the end of the study, how long it would take them to complete the reading if they were going to be tested and whether they would have any background knowledge on the subject we were going to study them learning.

Then it seemed meaningful to study how people learn under the pressure of a post-test and without it, considering that most methods of learning today involve a written test. While there are many different methods that classes use to teach students, a common form of learning is reading from a textbook or assigned readings. Thus we set the scope of our study to the reading behavior and patterns of students. We intended to measure the impact of prior knowledge of assessment on test scores and learning, and the amount of times it takes to read a text if subjects know they will be tested prior to starting the reading or not.

Our hypothesis is that students who

know that there will be a test prior to doing the reading will have longer reading times. In addition, when comparing the fixation maps of the two study groups, we will find that students with prior knowledge of the test will spend more time on new terms, definitions and numerical data, thus have a higher average score on the assessment test.

Background Research

There has been numerous studies on the past of the effects of assessment tests onstudying and learning. While we won't go into detail about each of the foundings of the studies we researched, we will mention two studies that played an important role in shaping our own.

"Prior Knowledge and Online Inquiry-Based Science Reading: Evidence From Eye Tracking" by Hsin Ning Jessie Ho, Meng-Jung Tsai, Ching-Yeh Wang and Chin-Chung Tsai is the first of these. The motivation of the study is to examine the role of prior knowledge in distributing visual attention on scientific information and data diagrams. For this study, participants are split into groups of low and high prior knowledge after pre-assessment, and presented the same reading on global warming. The total reading time, total fixation duration, total regression number and inter-scanning transitions between text and graphics are measured using eye trackers. In addition, fixation maps are made from the eye tracker data of all participants. The results of the study show that high prior knowledge participants spend more time on diagrams, have longer fixation durations and more inter-scanning transitions. They are also better able to integrate text and graphic information.

This study was valuable for us for two main reasons. After reading about the effects of prior knowledge on reading behavior, we realized the need for a group of participants with the same level of prior knowledge on the subject of our reading. To control prior knowledge, we chose a reading on a very specific branch of brain development to test our participants with, and we decided to exclude any undergraduate participants who were studying or had prior knowledge about biology, neuroscience, medicine and other natural sciences. Moreover, seeing the effective data analysis done on the eye tracking data in this study was helpful in choosing our own data analysis method. Since we are not measuring differences between the image and text data presented in our reading, we decided to measure total reading time and make fixation maps in our own study.

The second study that helped us design this project was "Tracking Students' Eye Movements when Reading Learning Objects on Mobile Phones" by David Kabugo , Paul B. Muyinda , Fred. M. Masagazi , Anthony. M. Mugagga , Mathias. B. Mulumba. This study followed the experiences, observations and reflections of 68 Luganda language teacher trainees in using an emerging Educational Technology, eye trackers. The trainees picked excerpts from a Luganda language literary text and used the Tobii-T120 to collect gaze plot data on their students reading them. They presented their findings and their analysis of what the data means in a tabular format. This analysis was very helpful when we were doing our own analysis of the gaze plot data. For example, in this study trainees observed that students fixated more at the first segment of the reading than the subsequent segments. Their interpretation of this finding was that "when cognitive load increases as a result of reading and engaging with additional segments of a text, individual attention and fixation at subsequent segments of the text diminishes" (Kabugo, 60). This list of interpretations helped us shape our own analysis of our fixation maps.

In addition to reading past studies of similar nature, we conducted background interviews with undergraduate students at Northwestern University to learn more about their studying patterns and preferences. 11 out of 12 of the students we interviewed confirmed

that reading is their primary method of learning. Students also mentioned that they are often required to write reading posts or complete equivalent homework assignments after their readings to prove absorption of material. This confirmed our assumptions that students who are assigned readings are tested in some way or another. 4 out of 12 students said that they experience test anxiety, which effects their performance on assessment tests despite the fact that they read carefully and learn the material. In order to eliminate the effect of test anxiety on our study results, we conducted a short pre-test to prospective participants of our study. This test consisted of three questions we asked our participants: to rank their test anxiety, general test performance and how often they read on a scale of 1 to 10. We eliminated all potential participants who ranked their test anxiety higher than 5 and those who ranked their average test performance lower than 5. We also attempted to choose participants who ranked how often they read similarly.

Research Questions

1. Do tests improve the amount of knowledge that is retained from reading?

2. What kind of knowledge do readings in preparation for tests help students retain?

3. Do tests cause students to focus on certain kinds of words, numbers, or other materials that students reading just to read might not focus on?

4. Do students reading for a test spend a longer time reading than one who isn't?

Data Collection

We collected four different types of data for each participant in our study: total reading time, eye movement on the page, a video of the participant's face during reading and their assessment test score. We told participants prior to the study that they had as much time as they want to complete the reading. This ensured that every participant read at their own pace and absorbed the information in the same way they would when reading for a class. Then we simply measured the total reading time using a stopwatch.

In our first wave of data collection, we used the Eye Tribe eye tracker to collect gaze data on 10 participants. However, we faced some difficulty getting accurate gaze fixation data for the entirety of the pages. The Eye Tribe tracker would work fine for a little while, but start to only gather gaze from a certain portion of the screen. Thus we decided to look into other eye trackers. In our second wave of data collection we switched to the Tobii Eye Tracker 4C. We used the Tobii to monitor where the user is centering their attention and for how long. This data gave us an idea as to what students are focusing on if they're preparing for a test vs reading for their own benefit. During the test, the Tobii eye tracker was placed at the foot of the screen of the laptop the participants were reading from. We first calibrated the eye tracker for each participant and then started timing their reading time. Participants were not allowed to scroll through the reading in order to keep coordinates of the page constant, so they indicated switching between pages by clicking on the page. The data collected thus had the text "User Clicked." printed for the moments where participants were moving on to the next page of the reading.

For the assessment test score, we prepared a simple ten question assessment test that included four multiple choice questions, two fill in the blank questions, one true or false question and three open ended questions. By including different types of questions, we assessed both recall and comprehension immediately after the reading. The assessment test was worth a total of 12 points, so each participant received a score out of 12.

In addition to using the Tobii Eye Tracker, we recorded the participants using Open Broadcast Software in hopes of being able to get information about pupil dilation of head movement.

The Tobii Eye Tracker has an extensive

SDK. We wrote a C# project that made use of the SDK in order to gather the gaze data fixation in coordinates for the participants and then used a python script with pandas and matplotlib to plot and visualize the coordinates and map them to what the participants read. This python script was also used to clean the data and get analytics about how much time was spent looking at the actual text.

Data Pre-Processing and Analysis

The total reading time data collected and the assessment test scores were compiled in a spreadsheet. This spreadsheet is attached below. The Tobii data collected was in the form of a set of coordinates for each participant. We saved the coordinates for each participant as a text file to later use in our analysis. We later put the x and y coordinates of these coordinates in separate arrays to make our fixation maps. Every gaze coordinate that was not on the document was removed and factored into the calculation for how much time the participants spent looking at the the reading as opposed to something off screen or off the text.

To analyze the total reading time we first took the average reading time for the two study groups:

Prior Knowledge of Test Average Time = 3:09 minutes

No Prior Knowledge of Test Average Time = 3:05 minutes

We also calculated the average assessment test score for the two study groups:

Prior Knowledge of Test Average Test Score = 9.3 out of 12

No Prior Knowledge of Test Average Test Score = 5.5 out of 12

In addition to comparing the two groups together, we wanted to understand the relation between the total reading time and assessment test performance for all the participants.

Findings

We found that the average total reading time for the two groups were almost identical. We gave participants as much time as they wanted to complete the reading and our hypothesis was that having prior knowledge of the test would cause participants to spend more time on the reading in attempt to perform well on the test. However the data shows no such tendency. We can account this lack of a difference between the two groups to the different reading speeds of each participant. Participants who read more in general spent less time completing the reading than those who do not read as much. On the other hand, we found that there was a significant difference between the average assessment test scores of the two groups. The group with the prior knowledge of an assessment test performed much better than the group without, on average about 3.8 points or about 31% better on the test. This fits our initial hypothesis that prior knowledge of a test would help enhance performance.

When we examined the relation between the total reading time and the test scores of all participants, we did not find a significant correlation. Reading the text for a longer time did not mean that you performed better on the test. This makes sense since the largest factor affecting reading time is personal reading speed, and reading the text for longer does no necessarily mean that you are paying more attention.

When examining the fixation maps of all participants we noticed a few differences between those who were aware of the assessment after the reading and those who were not. One of the main differences had to do with how often participants examined images. Although there were a couple of participants who examined the images regardless of an assessment after the reading, those who were aware of the assessment spent more time looking at images if the images were pertinent to what could be tested, while

the other group had participants who spent no time looking at images that had testable content. On the contrary, participants who were unaware of the test actually spent more time looking at pictures, such as the fetus in Figures 3 and 4, that might interest them.

Another significant finding from the fixation maps is the amount of time looking at the actual text of the reading. Although both groups spent about the same time reading the text on average, the group who was aware of the assessment was much more focused on the text during that time. In the fixation maps for the group who was unaware, we were able to see many more fixation points completely off of the text, while those who were aware of the assessment had the majority of their gaze focused on the actual text. This could also factor into the test scores for the two groups, as the group who directed their focus on the material had higher test scores.

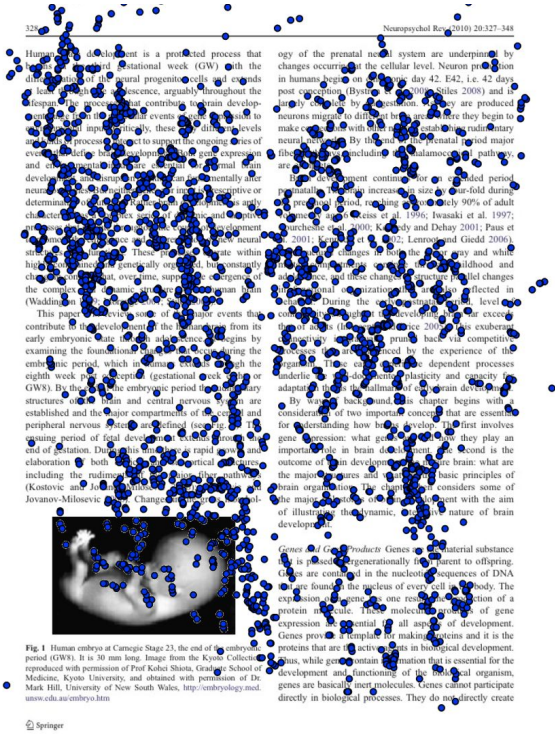

Figure 3: Participant was unaware of assessment. 65.4% of coordinates were focused on the reading.

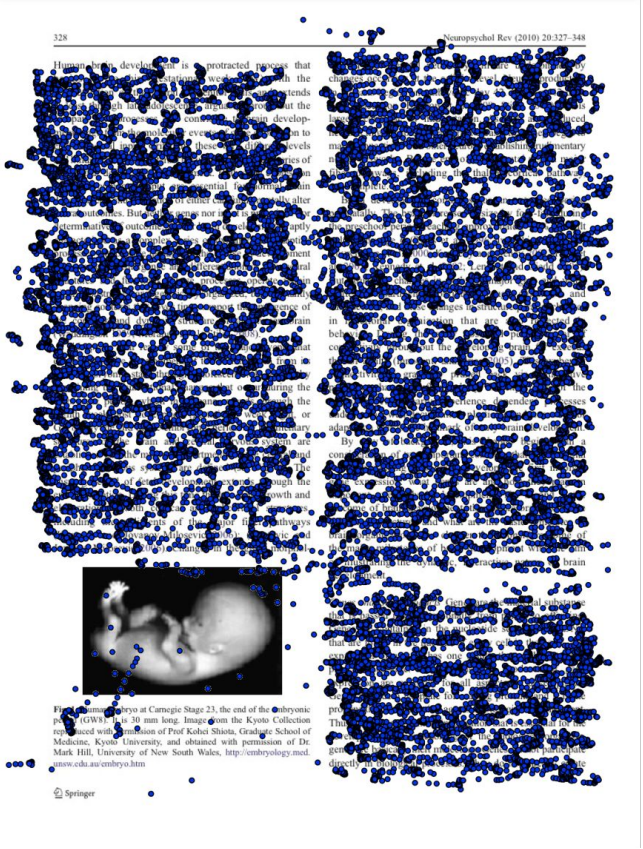

Figure 4: Participant was aware of assessment. 80.8% of coordinates were focused on the reading.

The fixation maps in figures 3 and 4 demonstrate that an assessment causes participants to focus on the reading at hand. The participant in figure 3 only had 65.4% of its coordinates focused on the reading, while the participant in Figure 4 had 80.8%.

Limitations

A limitation we faced was controlling the personal reading speeds of the participants. We could not guarantee that the participants all read at similar speeds normally, which made the measurements of total reading time less valuable. We initially thought to ask prospective participants to rank how fast they read in the pre-test as well, but we decided that the participants themselves would not be good judges of how fast they read. We attempted to control this variable by including the "How often do you read?" question in our pre-test. In the future, a more accurate

pre-test can be conducted prior to the study where prospective participants are asked to read two or three of the same passages. This way, their true reading speeds can be calculated, and only those with almost identical reading speeds can be selected for the study. Then the reading speeds would be controlled and a more accurate study of the effects of reading time on assessment test score can be done.

Another limitation was our participants' lack of personal investment in the results of the assessment test. When undergraduate students are reading to prepare for a test in real life, they are under the stress of wanting to perform well. Unfortunately, our participants do not feel this same sense of responsibility to perform well in our assessment test. This can have effects on their reading behavior like how carefully they read facts, how much time they spend on figures and so on. Prior to taking the test, many of our participants asked us if the score they get on the assessment test matters or counts for anything. This was a difficult question to answer because we did not want to affect their reading behavior, but knowing that the assessment test would not affect their life outside of our experiment altered their approach to the material.

Similarly, prior knowledge of our participants on the subject of the reading was also a limitation. We attempted to control this variable by selecting participants who were not studying or had prior experience past high school in biology, neuroscience, medicine and other natural sciences. Most of our participants were computer science or some sort of STEM major from a wide range of universities and ages. However, this does not entirely eliminate the effects of prior knowledge on the assessment test results. Two of our participants actually noted after the study that they somewhat remembered some of the content of the reading from their high school biology classes.

Lastly, the Tobii Eye Tracker presented some limitations. The eye tracker seemed to be much more accurate than the Eye Tribe in terms of recognizing exactly where your gaze was; however, we were unable to gather gaze coordinates as frequently. Because of this, some fixation maps didn't give us as much information as we would've liked. It is possible to change the frequency at which coordinates are picked up as we were able to find some examples in C++ that were able to do this, but we were unable to replicate it in C#. Therefore, for participants who read quickly, the fixation maps look as though some data points are missing.

Future Work

Future work for this study could improve on the testing conditions of our study. For example, to make user behavior more organic and thus the data collected more realistic, the Tobii Eye Tracker can be prepared so that the reading is scrollable. Moreover, the study group size can be expanded to include many more participants. In our first wave of the study, we tested 10 participants using the Eye Tribe Eye Tracker. In this second wave, we tested with another group of 10 participants, 5 in each study group. While we have data collected from a total of 20 subjects, this is still a fairly small sample size. In the future, this study can be expanded to include a hundred participants, or the population of a specific school of the university. The study could also include a second assessment test sometime after the initial data collection where the participants are given another assessment test on the same reading they did. This way, we can measure how well the participants retained the information they read.

While we collected video data of our participants, in the future we would like to use these videos to record pupil dilation data. Pupil dilation can give us information on participant engagement and shine more light into the focus difference between participants preparing for a test and the ones reading for leisure. This is something we can easily implement as we

are already able to correlate the start time of the reading in the video the timestamps collected with the gaze coordinates. This means we will know at what point in the reading a participant is at if we see dilation.

More variables can be measured in the future to offer more insight into the effects of prior knowledge of testing on reading and performance. For example, a brain activity device such as the Muse headband can be used to measure engagement throughout the reading. This would shine light onto what percentage of the total reading time the participants were focused on the reading, and how active their brain was when preparing for a test compared to reading for enjoyment.

Conclusion

To conclude, the study showed that prior knowledge of assessment tests makes students pay more attention to facts, terms and numerical data. Therefore students who have prior knowledge of post-tests perform better in these assessments. Our data showed that our hypothesis was partially correct. Students who knew that there will be a test prior to doing the reading did not have longer reading times; however their fixation maps showed that they spend more time on new terms, definitions and numerical data and consequently perform better in the assessment tests. It must be noted that our findings only show that prior knowledge of assessment leads to better performance in these assessments, not necessarily that this leads to better learning. Undergraduates students are used to taking tests. Perhaps they alter their reading behavior to prepare for test questions, paying more attention to these new term definitions and numerical data that they quickly forget after the assessment is complete. Whereas students reading for leisure might absorb information that they find interesting from the reading, instead of focusing on memorizing facts that they might later on be tested on. Future work must be done to answer these new questions accurately.

References

1. Ho, Hsin Ning Jessie, et al. "Prior Knowledge And Online Inquiry-Based Science Reading: Evidence From Eye Tracking." International Journal of Science and Mathematics Education, vol. 12, no. 3, 2013, pp. 525–554.
2. Kabugo, David, et al. Tracking Students' Eye-Movements When Reading Learning Objects on Mobile Phones: A Discourse Analysis of Luganda Language Teacher-Trainees' Reflective Observations. Manhattan Institute for Policy Research. 52 Vanderbilt Avenue, New York, NY 10017.
3. Tara Brady, Camille Salas, Ayah Nuriddin, Walter Rodgers & Mega Subramaniam (2014) MakeAbility: Creating Accessible Makerspace Events in a Public Library, Public Library Quarterly, 33:4, p. 331, DOI: 10.1080/01616846.2014.970425