In the Eyes of the Beholder: How Experience Can Influence Perceptions of Interest and Engagement

Aiya Saad

University of Michigan

Mentors: Dr. Kevin Miller and Dr. Kai Cortina

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Bachelor of Arts in Psychology from the University of Michigan 2018

Abstract

The curse of knowledge is a cognitive bias where an individual unknowingly assumes that an individual they are conversing with has the same background of information to understand. Previous research has shown that this bias is present everywhere including adults and younger children. I question whether graduate student instructors fall victim to the curse of knowledge and if they will overestimate or underestimate undergraduate students' ratings of instruction as well as level of engagement. With the use of continuous data collection via joystick, I attempt to obtain data to uncover how this information imbalance between graduate student instructors and undergraduate students can affect classroom settings. I hypothesize graduate student instructors will fall victim to the curse of knowledge and because of this, they will overestimate undergraduate students' ratings. If my hypothesis holds true, it can suggest that this information imbalance can effect classroom settings for both parties involved.

Key words: Knowledge, Graduate, Undergraduate, Lecture

In the Eyes of the Beholder: How Experience Can Influence Perceptions of Interest and Engagement

On a daily basis, individuals engage in conversations in which the exchange of various degrees of knowledge takes place. Achieving what is considered a "good" discussion comes with an understanding of both parties. Unfortunately, this is not the case, and there is often an intellectual imbalance between parties. The curse of knowledge demonstrates how individuals tend to assume others are just as informed as they are in a field or subject, and how it makes it challenging to interpret information undoubtedly. This curse is a cognitive bias enabling individuals to assume their peers receive information more accurately than they truly do. For example, a Chemistry professor who has obtained a Ph.D. in Chemistry may find it frustrating when a novice, for example, an undergraduate student, does not comprehend a particular topic. The students clean slate on this topic versus the professor's Ph.D. degree makes it difficult for the professor to cater to his pupil's confusion patiently. It is important to note that neither the student nor the professor is at fault for this lack of understanding, it is in fact, the works of the curse of knowledge. Awareness of this disparity was made possible through lab studies, and more research on this matter continues to expand our knowledge on this curse.

In the early nineties, doctoral candidate Elizabeth Newton depicted the phenomenon that we know today as the "curse of knowledge." Newton invited college students at Stanford University to participate in her experiment and accept one of two roles, tappers or listeners. Tappers were given a list of twenty-five well-known songs and were asked to tap out the rhythm of each song. When asked to predict how many songs the listeners would be able to guess correctly, tappers guessed fifty percent. However, when listeners were asked to identify the songs tapped by the tappers, listeners only guessed the tune correctly about three percent of the

time. When asked what they heard, listeners described the taps as being disconnected and confusing which made it difficult to make out the tune (Netwon,1990). Newton's experiment presented the difficult task of communication between the two groups. The small amount of information given to the tappers allowed them to have insider knowledge of the familiar tune in their mind while tapping it out. On the other hand, as a result of no information being given, the listener was only able to hear random tapping. This "curse of knowledge" emphasizes a problem for all: once one obtains bountiful knowledge on a particular subject, the skill to communicate with others on the matter becomes challenging because it is assumed the other party is just as familiar in the field. This belief, as exhibited in Newton's study, is problematic in a number of settings for communication and has influenced other researchers to expand on her findings and conduct studies on what Newton's experiment lacked.

A variation of the curse of knowledge can be seen in young children, ages three and under, have a hard time interpreting false beliefs. A well-known demonstration of this concerns a young character named Sally who places a box of chocolates into a box and then goes outside. Later, a different individual enters the room and places the box of chocolates into a basket. Adults and young children are asked, "when Sally returns to the room, where will she look for her box of chocolates?" Adults and children over the age of three answer by stating Sally will look for the box where she initially placed it. In contrast, children under the age of three assert she will search where it currently is, in the basket. This classic study has been repeated utilizing various methods, but the general findings remain the same. Numerous researchers perceive this poor performance as reflecting a qualitative difference in the ability to appreciate mental states. Others, however, argue that young children have a problem reasoning about false beliefs is due to more general factors such as memory loads, processing limitations, etc. and not necessarily

due to a lack of understanding. In an influential experiment, Birch and Bloom explore an innovative explanation for these developmental differences that are most consistent with the belief that young children's difficulties result from general factors. Birch and Bloom suggest that these false-belief tasks are cursed since the participant knows where the chocolate is, s/he will be biased to assume the character in the task knows too. The researchers hypothesize that even though, "the curse of knowledge for adults and older children is subtle, enabling them to succeed in false-belief tasks, it is greater for young children, and thus a source of difficulty in mental-state attribution" (Birch & Bloom, 2003). If this holds true, the curse of knowledge will be made clear in tasks that do not involve false-belief assessment, and should be more effective in younger children, who typically fail the task, versus older children, who usually pass the task.

To examine this hypothesis, researchers presented children between the ages of three and five with two sets of toys, one described as recognized by the experimenter's puppet friend, Percy, and the other as being unfamiliar to Percy. Experimenters informed the children that each toy had an object inside and were asked to consider whether or not Percy would know what was inside the toys. Children were randomly assigned to one of two groups; child-knowledgeable, these children were shown the contents of the toy, and child-ignorant, these children were not shown the contents of the toy. Experimenters predicted that children who were assigned to the child-knowledgeable condition would overestimate Percy's knowledge of the objects in the toy. On the other hand, researchers predicted that children assigned to the child-ignorant condition would not overestimate Percy's knowledge of the objects in the toy. Additionally, the researchers predicted that the inclination to overestimate would decrease with age. Results suggested that young children are especially sensitive to the curse-of-knowledge bias recognized in adults, which in turn, leads them to make mistakes in mental-state attribution.

In *Made to Stick: "Why Some Ideas Survive and Others Die"* (Heath & Heath,2003), authors Dan and Chip Heath propose why some ideas fail to stick by utilizing the curse of knowledge. They point out that, "Once we know something, we find it hard to imagine what it was like not to know it. Our knowledge has 'cursed' us. And it becomes difficult for us to share our knowledge with others, because we can't readily re-create our listener's state of mind" (Heath & Heath,2003). Throughout their dialogue, it is emphasized that society spends a lot of time preparing individuals to develop thoughtful answers but no time is invested in teaching people how to communicate these answers. Resolving this, suggest Dan and Chip, can be done by applying principles when communicating, allowing for a change in the way we communicate as a society. With the acronym S.U.C.C.E.S, Chip and Dan attribute each letter to a characteristic that can help make one's idea "stick." An idea needs to be simple and focused, unexpected and an attention grabber, concrete and understandable, credible and agreeable, emotional so people care enough about it, and ultimately, there needs to be a story behind it.

In *No Second Chance to Make a First Impression*, researchers question whether it is possible for instructors to recover from a bad first impression. It was hypothesized that first impressions may color student experience of instruction regardless of the lesson quality and these first impressions may provide valid evidence for instructional quality. Two studies were ran to investigate the correlation between first impressions and instruction quality compared to learning and evaluation of instruction among college students. For the first study, it was predicted that if a first impression has an effect on the evaluation of instruction, then there will be a strong effect of the introduction independent of the quality of later instruction. However, if first impressions generally accurately predict what follows, then the instructional quality should be the main predictor of learning and evaluation. Students were randomly assigned to a video

which consisted of a male instructor lecturing on topography which had either good first impression and good quality of instruction, good first impression and bad quality of instruction, bad first impression and good quality of instruction, or bad first impression and bad quality of instruction. Students were then asked to rate the instructor on a scale of 1-10 on fourteen dimensions, such as confidence and honesty. The second study sought out to test the robustness of the results of the first study when the lecturer and the topic were changed. Study two consisted of the same four conditions as the first study but the lecturer was now female and the topic was international comparisons. The researchers found that, "Quality of instruction is the strongest determinant of student factual and conceptual learning" (Samudra, Min, Cortina, & Miller, 2016). In general, the results from study one were replicated in study two. Instructional quality had a stronger effect on both learning and teacher evaluations, than first impressions did.

The curse of knowledge is present in classroom settings and in the current study I will be extending this research to test how well graduate student instructors actually know their students. I question whether graduate student instructors fall victim to the curse of knowledge and if they will overestimate or underestimate undergraduate students' ratings of instruction as well as level of engagement. Using a continuous monitoring system to track students' perceptions on quality of instruction and level of engagement. I will test whether graduate students rate instructors and their levels of engagement in a similar fashion. I hypothesize graduate student instructors will fall victim to the curse of knowledge and because of this, they will overestimate undergraduate students' ratings.

Method

Participants

Participants of the study were from the University of Michigan's Ann Arbor campus. Participants included undergraduate (n=41) and graduate students (n=12). Undergraduate students were enrolled in an introductory course in psychology and participated in this study for course credit. Graduate students were part of the University's Combined Program in Education and Psychology and were compensated for their time.

Measures

A lecture was given on international comparisons in education. The instructor gave a good first impression. This consisted of her introducing herself to the audience, as well as building ethos by explaining why she is passionate about the topic of international comparisons. She followed her introduction with a good lecture. She was very professional in her tone, articulate in her speech, well versed in her topic, and maintained eye contact with the viewer throughout the video.

During the course of their viewing time, students used a continuous monitoring system to track perceptions on quality of instruction and level of engagement. This consisted of a physical joystick which was to be held by the viewer for the duration of the video. On the screen, there was a pointer calibrated to the joystick overlaying a graph which was labeled with four different response types (Lizdek, Sadler, Woody, Ethier, & Malet, 2012). In order to gage level of engagement, I used the y-axis to measure two response types. The maximum of the y-axis was labeled 'learning a lot', and this corresponded with the viewer pushing the joystick north. The minimum of the y-axis was labeled 'learning very little' and this corresponded with the viewer pushing the joystick south. In order to measure instruction quality, I used the x-axis to measure two response types. The maximum of the x-axis was labeled 'excellent instruction', and this corresponded with the viewer pushing the joystick east. The minimum of the x-axis was labeled

'poor instruction', and this corresponded with the viewer pushing the joystick west. *Figure 1* illustrates the joystick model. The software continuously tracked the movement of the joystick and registered the data every half second for the duration of the video.

Procedure

The experiment began with students signing consent forms agreeing to the collection and publication of this data. This was followed with collecting the data of the undergraduate students. Prior to data collection, students were directed to respond to a Qualtrics background questionnaire which asked for questions regarding their age, level of education, as well as their class standing. This was followed by instructions being read by research assistants to the participants. The study asked the participants to watch and continuously rate a lecture video while imagining themselves in a real classroom setting. The instructions prompted the viewer to rate the video in response to two elements. First, how interesting they found the lecture, and second how they felt about the instructor, as well as the instructor's method of presenting. Viewers were instructed to keep their hand on the joystick for the duration of the video, and push the joystick continuously providing a rating on the two dimensions. This was followed by instructions on how to use the joystick to respond to the lecture. The joystick instructions were read, followed by research assistants performing joystick use for the participants. Participants then began the task and watched and rated the video for 1219 seconds. Once the participant finished, research assistants granted them credit.

Once all undergraduate student data was collected, I then began data collection of graduate students. The experiment began once the graduate students read and signed the consent form. Once consent was obtained, students then began answering background information administered online through Qualtrics. Questions included asking students for their age and how

many semesters they have served as a graduate student instructor for. Once background information was collected, research assistants read the instructions to the participants. Participants were informed that undergraduates enrolled in an introductory psychology course on campus had participated in the study beforehand. Graduate students were then asked to rate the video as they watched it, just as the undergraduates did, but responding as they would think the typical undergraduate student in an introductory to psychology course. Participants were informed that they will be rating the video on two things: 1) how much they feel like they are learning, and 2) how they feel about the instructor and the instruction itself. Again, viewers were instructed on the joystick and were directed to keep their hand on the joystick for the duration of the video. Viewers watched and rated the same video which ran for 1219 seconds and once the data collection was completed, participants were compensated twenty-dollars for their time. *Figure 2* shows what was on the screen that was provided to the participants.

Results

The level of engagement was measured for the undergraduate and graduate students with the joystick model between a scale of -1000 to a 1000 where -1000 meant completely unengaged, zero meant neutral, and 1000 meant completely engaged. *Figure 3* is a line graph illustrating undergraduate students' ratings of level of engagement. For plotting purposes, the graph represents twelve participants in the study rather than the forty-one in the original test. The twelve were chosen at random. The Y-axis shows the student's levels of engagement during different times of the lecture. The X-axis measures the time of the study in seconds. The figure shows that overall, the participants were engaged during the online video lecture, however, one participant, participant 12, had fluctuated levels of engagement. The quality of instruction was measured for the undergraduate and graduate students with the joystick model between a scale of

-1000 to a 1000 where -1000 meant poor quality of instruction, zero meant neutral, and 1000 signified excellent quality of instruction. *Figure 4* is a line graph illustrating undergraduate students' evaluation of the lecturer. The same twelve undergraduate participants that were randomly chosen for measurement of level of engagement was also test for quality of instruction. The Y-axis shows the students' ratings of the instructional quality and the X-axis depicts the time the data was collected in seconds. The figure, as well, shows that the participants rated the quality of instruction similarly with some variation but participant 12 shows a lot of variation in their ratings.

The same measurement scale was used for graduate students as it was for undergraduate student for both dimensions of level of engagement and instructional quality. *Figure 5* is a line graph illustrating graduate students' ratings of levels of engagement. The graph represents all twelve graduate student participants. The Y-axis shows the level of engagement and the X-axis shows the time measured in seconds the data was collected. Participants had relatively similar levels of engagement, excluding participant 6 who had inconsistent results. *Figure 6* depicts how these graduate students evaluated the lecturer. Similar to *Figure 4*, the Y axis shows the evaluation of the instructor and the X axis shows the time in seconds that the data was collected. However, unlike *Figure 4*, this graph shows much more variation regarding the evaluation of instruction throughout the data collection.

Figure 7 is a line graph plotting the two populations of students comparing the average instructional quality per half second. Undergraduate students are represented by a red line while graduate students are represented by a blue line. The Y-axis shows the evaluation of instruction against the X-axis, the time the data was collected in seconds. When comparing the two groups, undergraduates consistently rate instructional quality lower than graduate students. *Figure 8* is a

line graph plotting both populations of students comparing the average level of engagement per half second. Undergraduate students are represented by a red line and graduate students are represented by a blue line. The Y-axis is labeled as level of engagement and the X-axis is time in seconds which the data was collected. *Figure 9* represents a correlation utilizing the average scores of evaluation of instructor between undergraduate and graduate students. Each point represents the average time for that half second time interval for graduate and undergraduate students. The correlation coefficient is -0.0147 thus there was not a strong relationship between the two groups. *Figure 10* represents a correlation utilizing the average scores of level of engagement between undergraduate and graduate students. Each point represents the average time for that half second time interval for graduate and undergraduate students. The correlation coefficient is 0.7489 thus indicating that there is a strong relationship between the two groups.

Discussion

The level of engagement for undergraduate students maintains an average rating. However, one participant exhibits an abnormal level of engagement that deviates from the rest of the group. Even when compared to the rest the sample size (n=41) this undergraduate was the only participant to exhibit this pattern. This participant is considered an outlier in the study which could impact the validity of the data. This could possibly be explained socially by how serious the participants were taking the study. For example, a participant committed to the study will take caution in accurately measuring their own level of engagement while a participant who is not fully committed to the study could very easily skew the data by not accurately recording his/her level of engagement. Similarly, in the graph that measured undergraduates' rating of the instruction quality this same pattern is found, which can also be explained by the varying levels of interest in the study by the participants considering how the outlier is the same participant as

in the level of engagement graph. It is important to note that how participants measure their level of engagement in the study is subjective and that one participant's level of interest may be measured differently than another participant who had the same level of interest. This subjectivity could have been better controlled if a rating system was implemented on the joystick itself such as a numerical system that indicated their level of engagement. If this system was implemented, the data may have been more consistent and accurate.

The graduate students perceived ratings for the undergraduates' level of engagement showed greater variation. For example, one graduate student measured their perceived level of undergraduate engagement more enthusiastically than the other graduate students and may not be an accurate representation of the overall ratings that the graduate students gave to the undergraduate students because of the subjectivity that joystick ratings were not controlled for. On average, the graduate students exhibited slightly more variation in their evaluation of instruction than in their level of engagement. This data supports the hypothesis about information imbalance that is present between two groups of people with varying levels of education.

The data suggests that there is no correlation, $r(2436)= -0.01$ ns, between graduates and undergraduate evaluation of instruction quality although the measurements taken by the graduate students were consistently higher than the undergraduates. A reason graduate students elevated level of instructional quality rating in comparison to the undergraduates can be explained by graduate students level of investment in this area of study. For example, an undergraduate student may not be committed to study psychology at an introductory level because of their lack of experience in the field and their varying levels of interest for the subject. On the other hand, the graduate participants were sampled from a pool of graduate students enrolled in the Combined Psychology and Education Program at the University. These participants have more

exposure to the material as well as an increased level of interest than that of an undergraduate student. However, the correlation between the level of engagement between the undergraduate and graduate students were positive, $r(2436)= 0.74$ $p<0.001$, which piques interest considering the lack of correlation in the levels of instruction quality. This can be explained because of the contents shown in the video included comparison of the United States' global educational ranking. Even though the demographic of the group was varying in nationality backgrounds (international vs. not international) all participants are students studying in the United States and therefore may have found some aspects of the video resonated with them similarly to their peers. Although the data shows proportional levels of engagement and instruction quality between the undergraduate and graduate student ratings, the graduate student ratings given for the undergraduates were consistently higher than the actual graduate student ratings. This overestimation supports the hypothesis of the presence of the curse of knowledge.

However, even with these results, it is important to keep in mind that the study was limited in several ways. The sample size for both groups of students was small n=41 (undergraduates) and n=12 (graduates). It is important to not generalize the results for all undergraduate and graduate students by increasing the sample size for both groups, it could be possible to similar results in other populations. Another limitation is the lecture is too short (about twenty-three minutes) and is a single lecture-based session to compare to a normal semester (about four months) with a lecturer. By conducting a semester-long experiment, where a lecture is shown to participants two or three times a week, the results will show different patterns because the students become more familiar with the lecturer. A third limitation in this study is the "good" video may not be good enough at other institutions. For example, if students from another institution were to complete this task, they may rate the instructional quality less than the students at this institution. A fourth

and very important limitation to be made aware of is the subjectivity of an individual's measurement of a parameter. For future research, it would be interesting to reverse the roles where graduate students rate the videos and undergraduate students are asked to rate the video how they think a graduate student would and test whether or not the curse of knowledge is still examined in that scenario. By utilizing the results, the data uncovered, we may be better able to guide graduate student instructors on overcoming this hurdle of information imbalance and enhancing the classroom experience for students and instructors alike.

Reference:

Birch, S. A., & Bloom, P. (2003). Children Are Cursed. *Psychological Science,14*(3), 283-286. doi:10.1111/1467-9280.03436

Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition,77*(1). doi:10.1016/s0010-0277(00)00096-2

Heath, C., & Heath, D. (2010). *Made to stick: Why some ideas take hold and others come unstuck*. New York: Random House Books.

Leslie, A. M., & Polizzi, P. (1998). Inhibitory processing in the false belief task: Two conjectures. *Developmental Science,1*(2), 247-253. doi:10.1111/1467-7687.00038

Lizdek, I., Sadler, P., Woody, E., Ethier, N., & Malet, G. (2012). A Computer-Joystick Method for Coding Interpersonal Behavior Continuously Over Time. *Social Science Computer Review*, *30*(4), 513-521

Newton, L. “Overconfidence in the Communication of Intent: Heard and Unheard Melodies.” Unpublished doctoral dissertation (Stanford, CA: Stanford University, 1990).

Samudra, P. G., Min, I., Cortina, K. S., & Miller, K. F. (2016). No Second Chance to Make a First Impression: The “Thin-Slice” Effect on Instructor Ratings and Learning Outcomes in Higher Education. *Journal of Educational Measurement,53*(3), 313-331. doi:10.1111/jedm.12116

Figure 1-the joystick model that was used for this study.

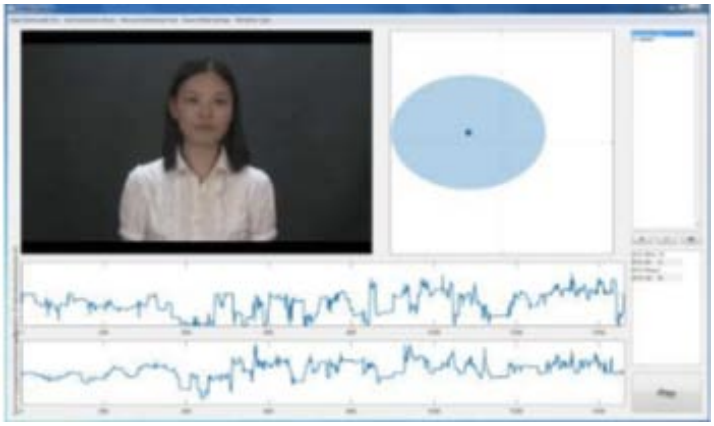

Figure 2- Example of a screen that was provided to students.

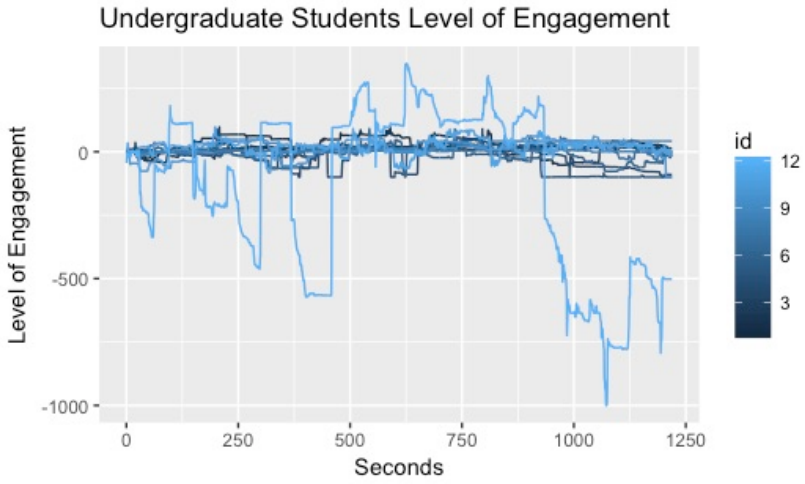

Figure 3-Line graph of Undergraduate student ratings of level of engagement against time. Figure depicts that the participants were engaged during the online video lecture, however, one participant, participant 12, had fluctuated levels of engagement.

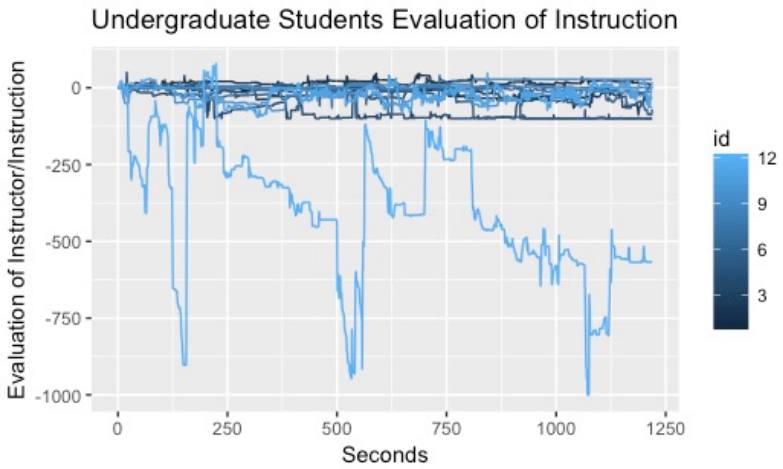

Figure 4- A line graph illustrating undergraduate students' evaluation of the lecturer. The figure shows that the participants rated the quality of instruction similarly with some variation but participant 12 shows a lot of variation in their ratings.

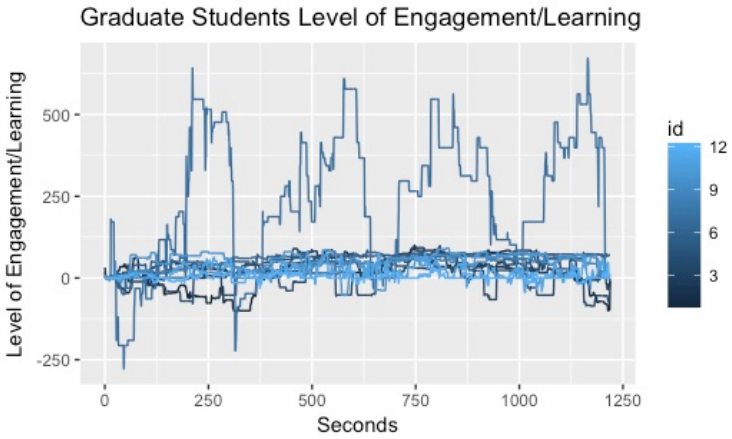

Figure 5- A line graph illustrating graduate students' ratings of levels of engagement. Participants had relatively similar levels of engagement, excluding participant 6 who had inconsistent results.

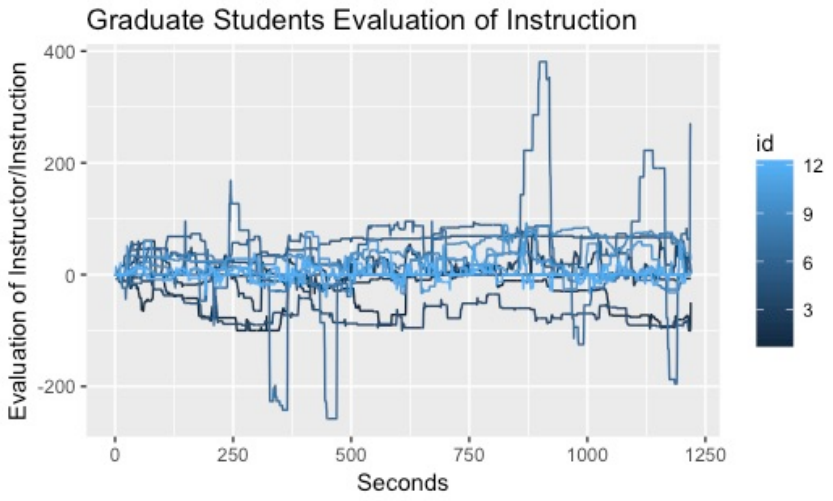

Figure 6- A line graph depicting how graduate students evaluated the lecturer. The graph shows much more variation regarding the evaluation of instruction throughout the data collection.

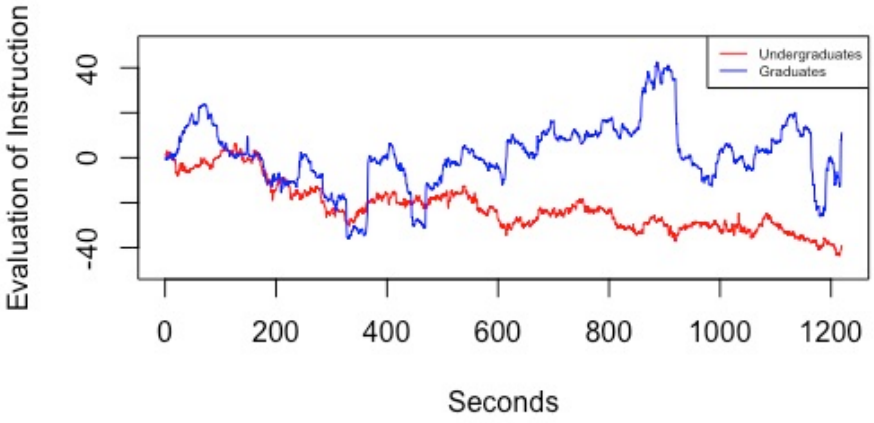

Figure 7- A line graph plotting the two populations of students comparing the average instructional quality per half second. Undergraduates consistently rate instructional quality lower than graduate students.

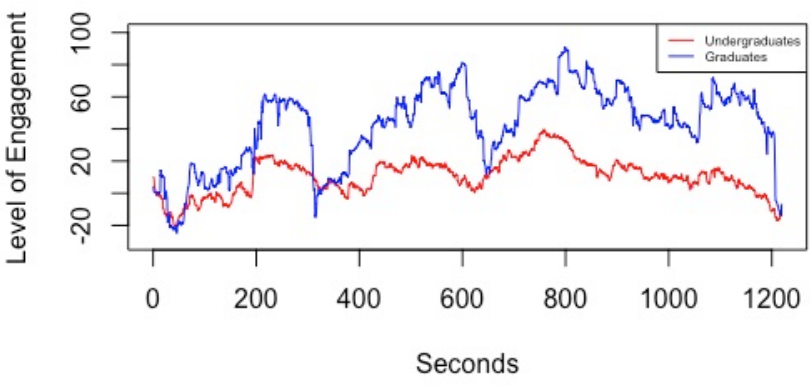

Figure 8- A line graph plotting both populations of students comparing the average level of engagement per half second. Graduate students consistently rate level of engagement higher than Undergraduates.

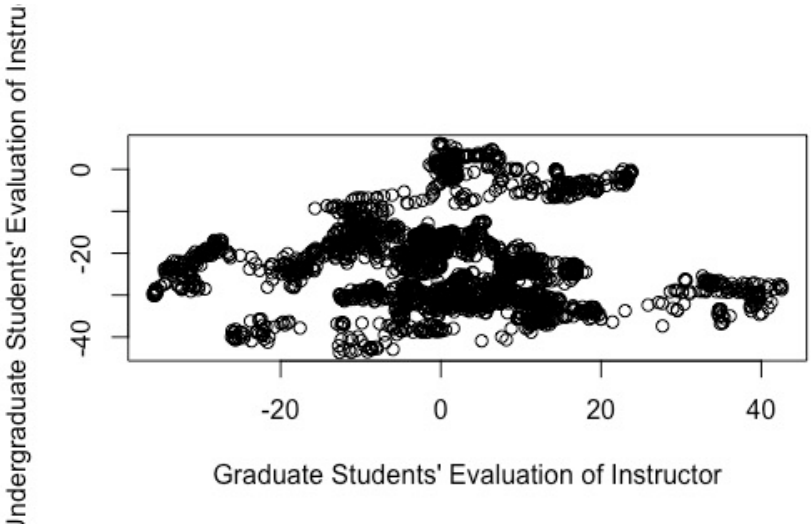

Figure 9- A correlation utilizing the average scores of evaluation of instructor between undergraduate and graduate students, r(2436)= -0.01 ns

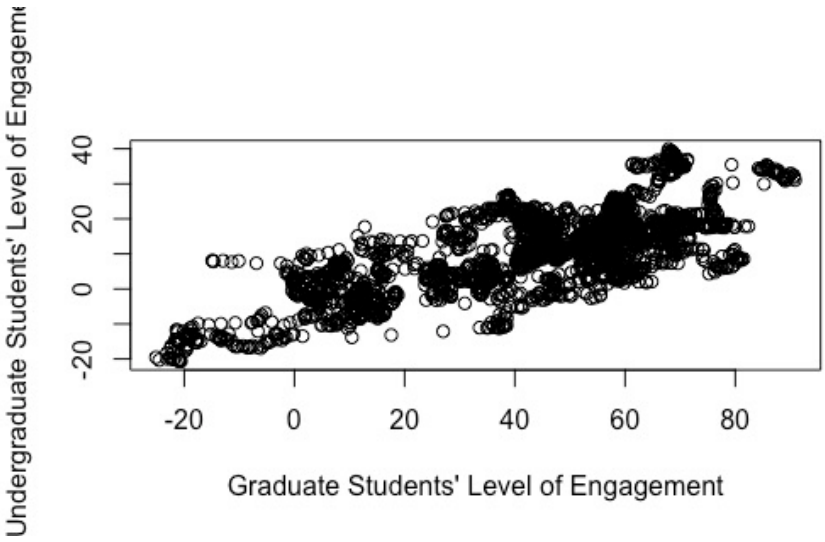

Figure 10- , A correlation utilizing the average scores of level of engagement between undergraduate and graduate students, $r(2436)= 0.74$ $p<0.001$