Big Ideas from Big Data: Data-Driven Educational Ecosystem

Stephanie Olson, Thao Ha, Ioulia Rytikova, PhD., Mihai Boicu, PhD.

ABSTRACT

Big data analytics in learning environments is gaining popularity due to the enormous amounts of data being generated by universities and other learning institutions through the use of web-based learning management systems such as Emodo, Moodle, and Blackboard. Virtual learning environments are being embraced by educational institutions at a rapid rate, pushing previously untracked classroom activity to timestamped and monitored systems capable of monitoring all student, course, and instructor activity. The rise of web-based virtual learning environments creates a new opportunity for data miners and analysts to observe trends and apply information to develop the most effective instructional techniques for the new virtual classroom environment.

One technique PLAIT researchers are analyzing is the spacing effect in the learning process. The spacing effect is the phenomenon where repetitive learning over long periods of time is believed to be more effective in the human learning experience than short-term rapid iteration. Many studies have shown that when new material is introduced to students, they exhibit increased retrieval and retention when the new material is spaced out over time, compared to an intensive massed practice of the material. However, the empirical data supporting the validity of the spacing effect globally in college level courses is insufficient. Research that can produce empirical data to determine the effectiveness of spaced learning is important to the learning sciences to assist in the creation of curriculums that optimize the human learning experience (for instance, by interleaving various topics at increasing levels of complexity, instead of discussing the topics one by one as in the traditional curriculum development). The PLAIT laboratory team is testing the effectiveness of spaced learning by comparing various aspects of spaced learning courses against traditionally structured ones.

Preliminary testing with R, Rstudio, and Excel is being performed on a dataset composed of 507 students, enrolled in 15 sections of the same course (IT214 Database Fundamentals), held between January 20 – December 19, 2015, at George Mason University. Statistical computations were performed on graded class components which include—but not limited to—exams, assignments, and labs to determine if significant findings in support of the spacing effect theory were present. These findings will contribute to see if the spacing theory is the most effective means of instruction for learning institutions.

VISION

The analysis presented in this poster is the beginning of a larger project which will help educators to better understand their students.

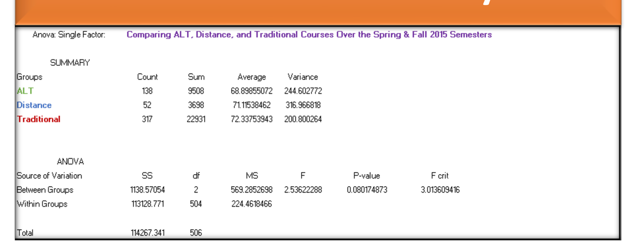
Researchers at the Personalized Learning in Applied Information Technology Laboratory (PLAIT) are working together to develop an educational ecosystem capable of analyzing all facets of the learning experience by examining all components and techniques in the learning environment. This will allow educators to improve instructional techniques by providing students with a personalized learning experience based on their habits, ability, and aptitude.

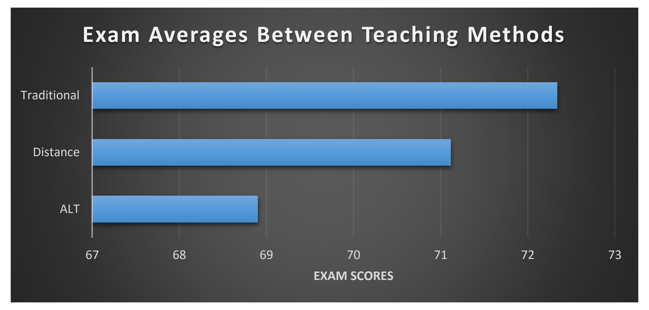
The goal of the PLAIT research laboratory is to incorporate and analyze the vast amount of Big Data generated from the virtual learning environment at George Mason University that is currently ignored. This growing data set will be utilized to examine student habits and instructional techniques in order to improve the effectiveness of the learning experience at George Mason University. The end result will be greater student academic achievement through the faculty's use of the most effective instructional techniques.

MATERIALS AND METHODS

The analysis represented in this poster focuses on one area of interest: that when utilized, spaced learning is the most effective teaching technique currently in practice at George Mason University. Preliminary data analysis was conducted using scores for the first exam given to all IT 214 students. The sample size consisted of 507 students, in 15 classes, during the Spring and Fall 2015 semesters. All students were enrolled in a course that was structured under one of three functional methodologies: Active Learning (ALT), Distance Learning, or Traditional in-person lectures.

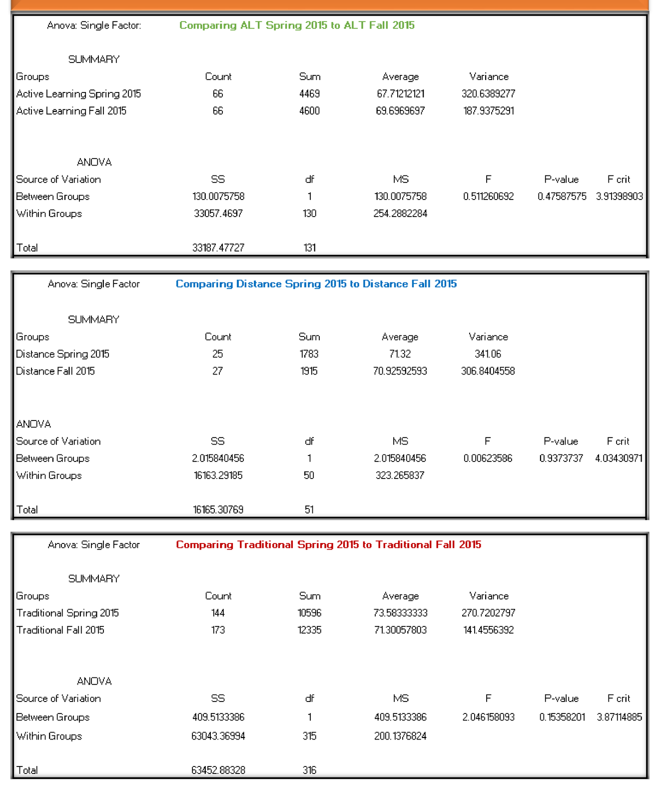
Course Overview Analysis



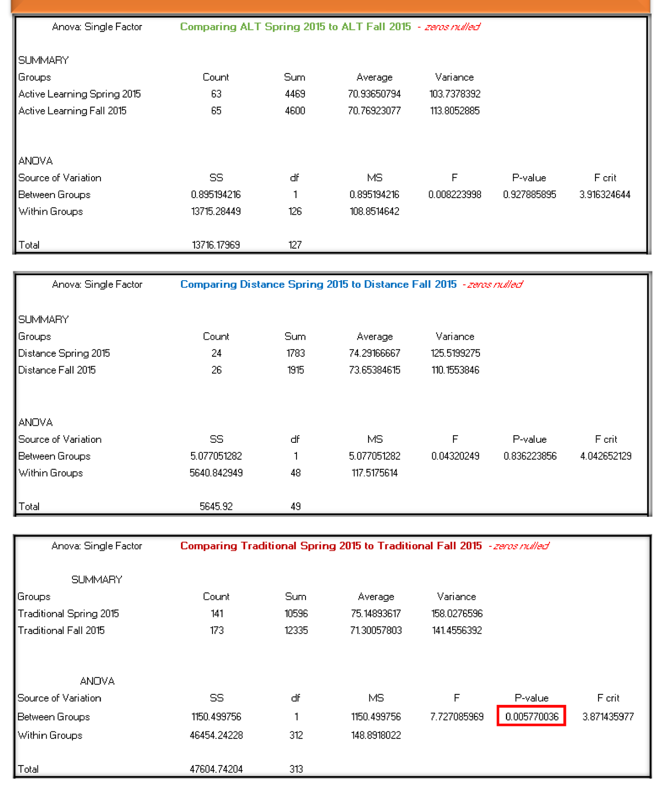


One-way ANOVA hypothesis test was performed to examine the difference in the mean of exam scores for students enrolled in the same course (IT 214), taught under three different methodologies: active learning, distance, and traditional in-person lecture courses. First, hypothesis testing was conducted to compare each course type against each other to determine if there was any noticeable difference between them.

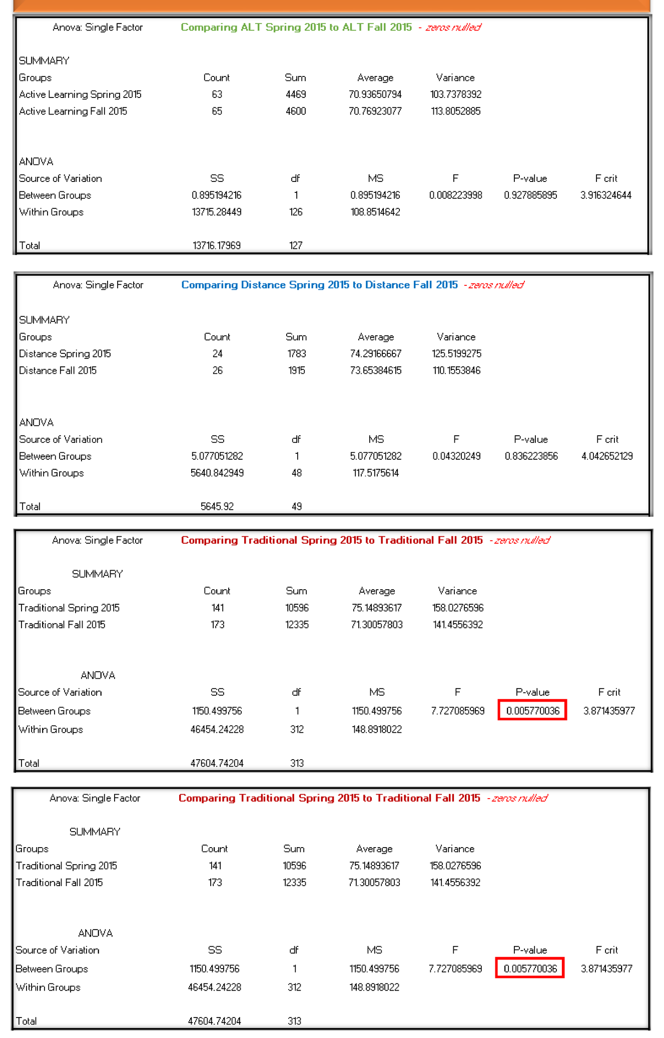
Course Type Analysis

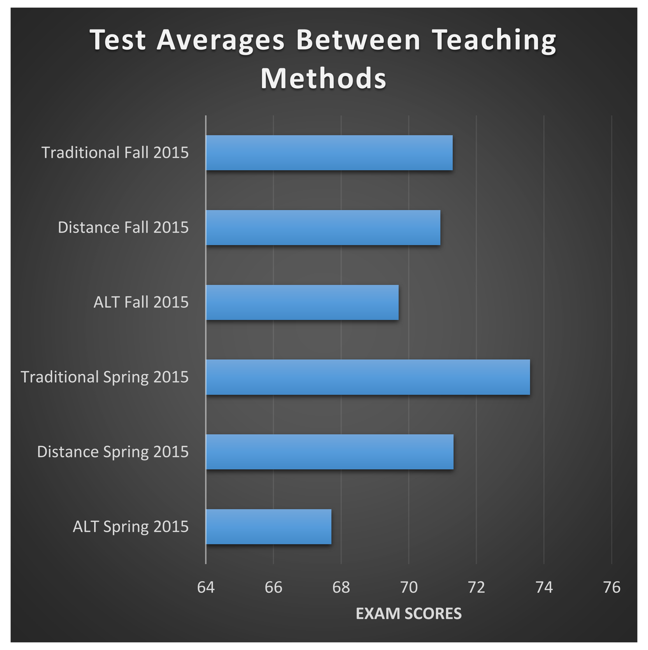


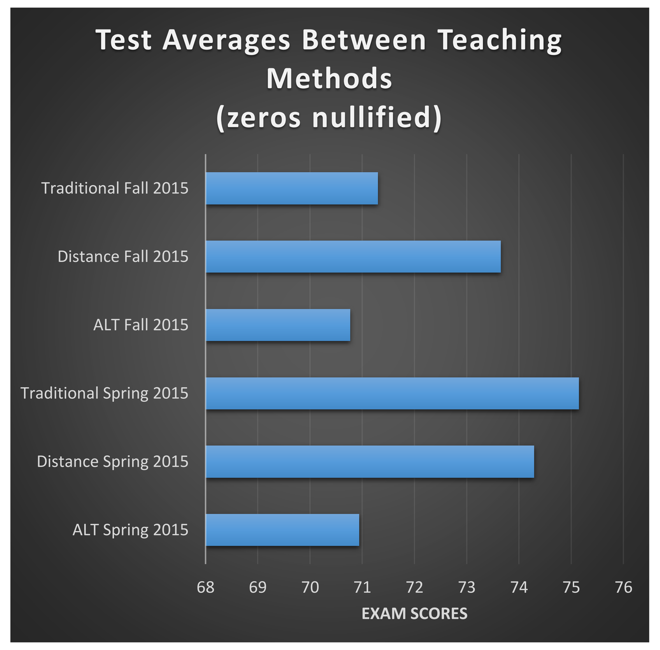
Experimental Analysis

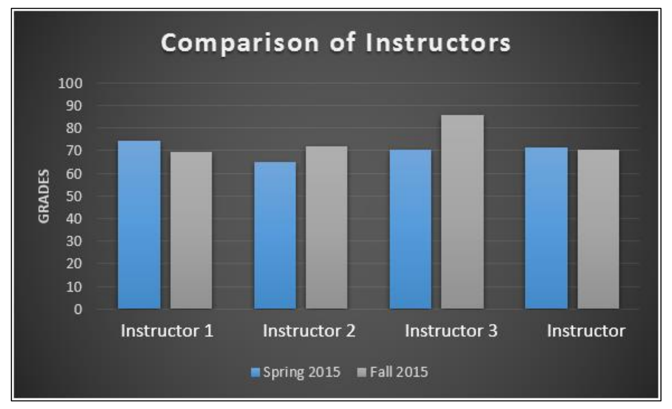


Instructor Analysis









No significant findings were found from the previous testing, so ANOVA testing was performed for the three course types in which Spring and Fall semesters were compared against each other. During analysis an inconsistency between the course types was observed due to a varying number of students who did not take the exam—resulting in a score of zero. Since this exam was the measure of performance in the class, only the test scores of the students who completed the exam were used and the zeros—from students who did not take the test—were converted to null values, and the test was conducted again. These results showed significant differences between the average exam scores of Spring and Fall traditional courses. The Spring 2015 average score was significantly higher, and a P-value of 0.00577 was produced. Due to the significant difference in the test scores between the Spring and Fall traditional courses, ANOVA tests were conducted again on traditional course instructors comparing their Spring classes to their Fall classes. This was performed to determine if there was a correlation between any particular instructor and the data anomaly of their students' exam scores.

CONCLUSIONS

Our current data set indicates that traditional in-person lectures produce better exam scores than the distance courses, and that the alternative learning environment is the least effective teaching method. Further, the data indicates an adverse effect on student exam scores when spacing techniques are implemented into the curriculum.

Tests on the data showed anomalies in the Spring 2015 exam scores for two traditional course instructors, which may have affected our initial testing results. The cause of these anomalies is not evident at this time and will require further analysis on a variety of other variables to determine if the findings are an accurate representation of the effects of space testing.

Additional research is required to determine the impact of the spacing effect, and it will be conducted using IT 214 course data over the past 10 years—providing an additional one million data points. Additional variables will also be analyzed to determine their impact on the learning experience in each of the aforementioned learning environments.

The inclusion of a much larger data set will allow PLAIT researchers to perform more meaningful analysis and afford us the opportunity to compare other aspects of the complex educational system in order to provide faculty and students with the best learning tools and methods available. Our current study utilized a data set which produced interesting results, but was too small and limited to provide us with an accurate overview of the impact of the methods and techniques currently in practice. Hopefully with the inclusion of these additional data sets (and more Big Data), our study will have enough information to produce significant and unwavering results.

REFERENCES

Brown, P. C., Roediger III, H. L., & McDaniel, M., A. (2014). Make it stick. *The Science of Successful Learning*. The Belknap Press of Harvard University Press, Cambridge MA.

McDaniel, M.A. (2012) Put the sprint in knowledge training: Training with spacing, retrieval and INTerleaving, . In *A.F. Healy & L.E. BourneJr (eds) Training Cognition: Optimizing Efficiency, Durability and Generalizability* (267-286).

I. Rytikova, M. Boicu. A Methodology for Personalized Competency-based Learning in Undergraduate Courses. In *Proceedings of the 15th Annual Conference on Information Technology Education and the 3rd Annual Conference on Research in Information Technology*, pp. 81-86, Atlanta, Georgia, October 15 -18, 2014

CONTACT

Stephanie Olson, Director of Communications/Graduate Researcher PLAIT Research Laboratory
Email: solson7@gmu.edu

Thao Ha, Graduate Researcher PLAIT Research Laboratory
Email: tha7@gmu.edu

Ioulia Rytikova, PhD., Co-Director PLAIT Research Laboratory
Email: irytikov@gmu.edu
Website: http://mason.gmu.edu/~irytikov

Mihai Boicu, PhD., Co-Director PLAIT Research Laboratory
Email: mboicu@gmu.edu
Website: http://mason.gmu.edu/~mboicu