

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.





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	Туре	e Anal	ysis		
Anova: Single Factor:	Comparing ALT 9	Spring 2015 to a	ALT Fall 2015			
SUMMARY	Count	C	A	Verinner		
Groups	Lount	Sum	Average	Variance		
Active Learning Spring 2015	66	4469	67.71212121	320.6389277		
Active Learning Fall 2015	66	4600	69.6969697	187.9375291		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	130.0075758	1	130.0075758	0.511260692	0.47587575	3.91398903
Within Groups	33057.4697	130	254.2882284			
Total	33187.47727	131				
Anova: Single Factor	Comparing Distan	ce Spring 2015	i to Distance Fall 20)15		
SUMMARY	- ·	_				
Groups	Lount	Sum	Average	Variance		
Distance Spring 2015	25	1/83	/1.32	341.06		
Distance Fail 2015	27	0181	70.32532533	306.8404998		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	2.015840456	1	2.015840456	0.00623586	0.9373737	4.03430971
Within Groups	16163.29185	50	323.265837			
Total	16165.30769	51				
Anova: Single Factor	Comparing Traditi	onal Spring 20)15 to Traditional F	all 2015		
SUMMARY						
Groups	Count	Sum	Average	Variance		
Traditional Spring 2015	144	10596	73,58333333	270,7202797		
Traditional Fall 2015	173	12335	71.30057803	141.4556392		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	409.5133386	1	409.5133386	2.046158093	0.15358201	3.87114885
Within Groups	63043.36994	315	200.1376824			



Experimental Analysis							Instructor Analysis						
Anova: Single Factor	Comparing ALT	Spring 2015	to ALT Fall 2015	- zeros nulled			Anova: Single Factor	Comparing ALT	Spring 2015	to ALT Fall 2015	- zeros nulled		
SUMMARY Groups Active Learning Spring 2015 Active Learning Fall 2015	Count 63 65	Sum 4469 4600	Average 70.93650794 70.76923077	Variance 103.7378392 113.8052885			SUMMARY Groups Active Learning Spring 2015 Active Learning Fall 2015	Count 63 65	Sum 4469 4600	Average 70.93650794 70.76923077	Variance 103.7378392 113.8052885		
ANDVA Source of Variation Between Groups Within Groups	SS 0.895194216 13715.28449	df 1 126	MS 0.835194216 108.8514642	F 0.008223998	P-value 0.927885895	F crit 3.916324644	ANDVA Source of Variation Between Groups Within Groups	SS 0.895194216 13715.28449	df 1 126	MS 0.895194216 108.8514642	F 0.008223998	P-value 0.927885895	F crit 3.916324644
Total	13716.17969	127					Total	13716.17969	127				
Anova: Single Factor	Comparing Dist	ance Spring	2015 to Distance I	Fall 2015 - 214708	nulled		Anova: Single Factor	Comparing Dist	ance Spring	2015 to Distance F	Fall 2015 - 278003	s nulled	
SUMMARY Groups Distance Spring 2015 Distance Fall 2015	Count 24 26	Sum 1783 1915	Average 74.29166667 73.65384615	Variance 125.5199275 110.1553846			SUMMARY Groups Distance Spring 2015 Distance Fall 2015	Count 24 26	Sum 1783 1915	Average 74.29166667 73.65384615	Variance 125.5199275 110.1553846		
NDVA Source of Variation Setween Groups Within Groups	SS 5.077051282 5640.842949	df 1 48	MS 5.077051282 117.5175614	F 0.04320249	P-value 0.836223856	F crit 4.042652129	ANDVA Source of Variation Between Groups Within Groups	SS 5.077051282 5640.842949	df 1 48	MS 5.077051282 117.5175614	F 0.04320249	P-value 0.836223856	F crit 4.042652129
Total	5645.92	49					Total	5645.92	49				
Anova: Single Factor SUMMARY	Comparing Tra	ditional Sprir	ng 2015 to Traditio	onal Fall 2015 -	zeros nulled		Anova: Single Factor SUMMARY	Comparing Trac	ditional Sprin	g 2015 to Traditio	nal Fall 2015 -	- zeros nulled	
Groups Fraditional Spring 2015 Fraditional Fall 2015	Count 141 173	Sum 10596 12335	Average 75.14893617 71.30057803	Variance 158.0276596 141.4556392			Groups Traditional Spring 2015 Traditional Fall 2015	Count 141 173	Sum 10596 12335	Average 75.14893617 71.30057803	Variance 158.0276596 141.4556392		
ANOVA Source of Variation Between Groups Within Groups	SS 1150.499756 46454.24228	df 1 312	MS 1150.499756 148.8918022	F 7.727085969	P-value 0.005770036	F crit 3.871435977	ANOVA Source of Variation Between Groups Within Groups Total	SS 1150.499756 46454.24228 47504.74204	df 1 312 313	MS 1150.499756 148.8918022	F 7.727085969	P-value 0.005770036	F crit 3.871435977
Fotal	47604.74204	313					Total	47004.74204	515				
Test	Averag	es B Met	etwee hods	n Tea	ching		Anova: Single Factor SUMMARY Groups Traditional Spring 2015 Traditional Fall 2015	Comparing Trad Count 141 173	Sum 10596 12335	3 2015 to Traditio Average 75.14893617 71.30057803	nal Fall 2015 - Variance 158.0276596 141.4556392	zeros nulled	
Traditional Fall 2	(ze	ros r		a)			ANDVA Source of Variation Between Groups Within Groups Total	SS 1150.499756 46454.24228 47604.74204	df 1 312 313	MS 1150.499756 148.8918022	F 7.727085969	P-value 0.005770036	F crit 3.871435977
Distance Fall 2	015	_	_					-		£ 1			
ALT Fall 2	015	_					100	compari	son d	or instr	uctor	S	
Traditional Spring 2	015		_				80 70 20 60		-				
Distance Spring 2	015						00000000000000000000000000000000000000						
ALT Spring 2	015						20 10 0						
	68 6	9 70	71 7 EXAM	72 73 SCORES	74 7	75 76	Instru	ctor 1 l	nstructor pring 2015	r 2 In s 5 ■ Fall 2015	structor 3	Insti	ructor



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.

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.

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 15thAnnual* Conference on Information Technology Education and the 3rd Annual Conference on Research in Information Technology, pp. 81-86, Atlanta, Georgia, October 15 -18, 2014

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

Thao Ha, Graduate Researcher PLAIT Research Laboratory Email: <u>tha7@gmu.edu</u>

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

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





CONCLUSIONS

REFERENCES

CONTACT