Learning Analytics Community Exchange

LAK15 Case Study 2: Examining Learners' Cognitive Presence in Massive Open Online Courses (MOOCs)

Learning Analytics Review: LAK15-2

ISSN:2057-7494

By: Vovides, Y., Youmans, T., Arthur, P., Davis, D., Ayo, E., Pongsajapan, R., McWilliams, M. and Kruse, A.

Published: 18th March 2015

Keywords: learning analytics, learning design, cognitive presence

This study examines a corpus of 4,825 discussion forum posts from 495 participants in a GeorgetownX MOOC on globalization for insight into the cognitive presence of learners and its implications for course performance. By analyzing the use of key terms linked to core course concepts as well as estimated level of language abstraction in the discussion forum, we examine the relationship between the results of this analysis, achievement, and video content engagement. By combining these varied analytics, we aim to get a better sense of learners' cognitive presence.

Contents

Executive Summary

This case study has been published as part of the practitioner track of the Learning Analytics and Knowledge conference LAK15, Scaling Up: Big Data to Big Impact, 16-20 March 2015[1].

Effective educational data mining has the potential to allow course instructors to make instructional interventions based on trends revealed through real-time learning analytics. With the end goal being to increase student achievement (their final grade, for example), this paper introduces linguistic trends in student data that have the potential to enable course designers and instructors to guide students toward better final grades. For example, we find that students who more frequently use words listed in the course's "Key Terms" glossary page in their discussion posts tend to earn a higher final score. Course designers can use this information by pointing and nudging students towards the Key Terms page and making it more accessible/visible throughout the course.

In our analysis of the students' cognitive presence in the GeorgetownX course, *Globalization's Winners and Losers*, we learned that students who exhibit high levels of language abstraction in their course notes and discussion posts tend to earn higher final grades. The same goes for students' video activity, higher video activity related to higher grades. While it may be less clear as to what design decisions can increase student language abstraction in their contributions, finding and sharing these trends with other course providers can lead us all to collaborate in asking new questions, starting new discourse, and exploring new course design solutions.

For all of these trends to be made available on a real-time basis while the course is running, participants and instructors would benefit from access to (and the invention of) a monitoring system/dashboard of sorts that serves to make both parties aware of the active learning sequence and all of its influencers. This learning analytics tool would enable students to potentially become more metacognitively aware and instructors to make targeted instructional interventions to maximize all students' cognitive presence.

1. Introduction

This case study examines a corpus of 4,825 discussion forum posts from 495 learners in a Georgetown University edX MOOC (GeorgetownX) on globalization to gain insight as to the relationship between learners' engagement with the course content presented in video format and course achievement (grade). This is done through an analysis of the use of key terms derived from core course concepts in discussion forum posts, as well as the level of language abstraction used in discussion forum posts. Language abstraction is a measurement of the depth with which a student is expressing conceptual understanding, with values of abstraction ranging from explicitly concrete to highly theoretic. To measure the level of language abstraction we used a lexicon developed by linguistic experts (Turney et al., 2011). Additionally, we examined the relationship between the key term use, language abstraction, activity with video material in the course, and achievement (course grade).

In defining cognitive presence, we turned to a framework known as the Community of Inquiry model (CoI). CoI is a conceptual framework, thus far used primarily in the context of traditional online courses, developed as a process model to "define, describe and measure the elements of a

[1] http://lak15.solaresearch.org/

collaborative and worthwhile educational experience" (Garrison, Anderson, & Archer, 2010, p. 6). Garrison et al. (2010) operationalized the model to include three core elements: social, cognitive, and teaching presences within the context of an online formal educational experience. The model emphasizes that it is the joining of the three presences that forms the optimal educational experience. While our current research specifically probes the cognitive element within this framework, we have designed the course to allow for future inquiry into the role of teaching and social presences as well.

In open online learning environments learner intent is a critical factor for retention and success, and through surveys we were able to account for intrinsic and extrinsic factors that may influence learners activity and performance in a MOOC environment. Koller et al. (2013) in their study of retention and intention identified three types of participants based on their pattern of activity in MOOCs – passive participants, active participants, and community contributors. What distinguishes these types is, according to Koller et al. (2013), learner intent. Koller et al. (2013, p. 63) explained:

> *"For MOOC retention metrics to be useful, they thus must be defined and interpreted with the learner's goals in mind. Passive lecture-watchers, for example, may go through an entire course without ever touching an assessment, yet they often derive substantial value from a MOOC without contributing to completion-based notions of retention."*

Two common reasons for active participation in open online learning environments are based on participants' personal interest or their need to develop specific competencies (Sheu, Lee, Bonk, & Kouu, 2013). With this in mind, we decided that the GeorgetownX learning design approach would cater to those students who would choose to be active participants. In other words, we designed the learning experience to support learners who wanted to engage with course content and assignment completion. However, we also approached the GeorgetownX learning design with modularity in mind so that participants could develop specific competencies by completing part of the course instead of needing to complete the entire course. For example, each topic/week of this GeorgetownX MOOC was in and of itself a complete module. To examine and further understand how participants engage with open learning environments so that we could improve upon our learning design strategy, we asked students to complete both a pre- and post-course survey. The pre-survey contains some demographic questions but mostly asks about the participants' intent and topic-level interest so that we can examine their cognitive presence in the course with their initial goal in mind. The post-course survey asks a series of questions to capture information about each of the three presences (cognitive, social and teaching) in the educational experience.

2. About This Case Study

2.1 Institutional Context

In 2012, Georgetown University announced a three-year Initiative on Technology-Enhanced Learning that had a dual focus: first, to improve teaching and learning by innovating with technology in mostly blended learning environments on our Washington, DC campuses; and second, to develop massive open online courses (MOOCs) in a partnership with edX to make aspects of a Georgetown education available to wider audiences around the world.

A little over two years into the initiative, we have conducted over 100 projects, engaged 174 faculty, and reached an estimated 5,200 students. We have launched eight MOOCs using the edX platform along with a custom-built platform that extends the edX platform capabilities. Our first MOOC, *Globalization*, is the focus of this case study. It ran for the first time in fall 2013 with approximately 35,000 students initially registered, and ran a second time in fall 2014 with approximately 10,000 students registered.

2.2 About the Case Study

Cognitive presence, a core element of the CoI model and the focus of this paper, is based on the Practical Inquiry (PI) model, which aligns with Dewey's ideas that experience and learning are intimately connected. The PI model is therefore grounded in Dewey's notion of reflective thinking as an active process of analysis and making judgments, resulting in a model that is developmental and includes four phases: triggering event, exploration, integration, and resolution (Boris & Hall, 2005). Figure 1 shows the four phases.

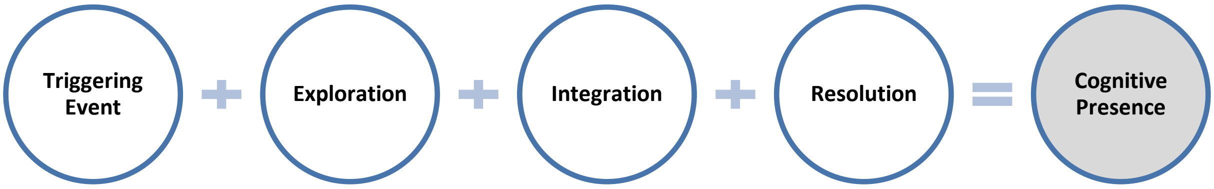

Figure 1: Four phases resulting in cognitive presence

In the *Globalization* course, we used the four phases to guide the design of the learning sequence to support learners' cognitive presence. For example, we used poll-based questioning prompts and key term video-based prompts to trigger exploration of the content and discussion-based questioning prompts to trigger integration of core concepts stemming from the content exploration. In addition, assessment questions were used in the sequence to capture the resolution or achievement. Table 1 maps each phase to the course design elements. It also shows the indicators being captured to provide evidence of learners' cognitive presence in a learning sequence.

Phase/Categories	Design Elements	Indicators
Triggering Event	• Questioning Prompts • Key Term Bubbles in Videos	• Student Lecture Notes (# of discussion posts made) • Polls (# of polls completed) • Video activity in relation to trigger bubbles
Exploration	• Student Lecture Notes	• Video activity overall • Key term use • Language Abstraction score
Integration	• Discussion Prompts • Country-based Activity Guides	• Key term use (in discussion and wiki) • Language Abstraction score (in discussion and wiki)
Resolution	• Assessment Questions • Poll Questions	• Problem activity (# of problems attempted) • Grade/points (# of points) • Polls (# of polls completed and # of times opinion changed)

Table 1: Linking design decisions to cognitive presence phases and indicators

Figure 2 shows an example of the use of a triggering event that encourages learners to take notes about key points from the videos, thus prompting their exploration of video content.

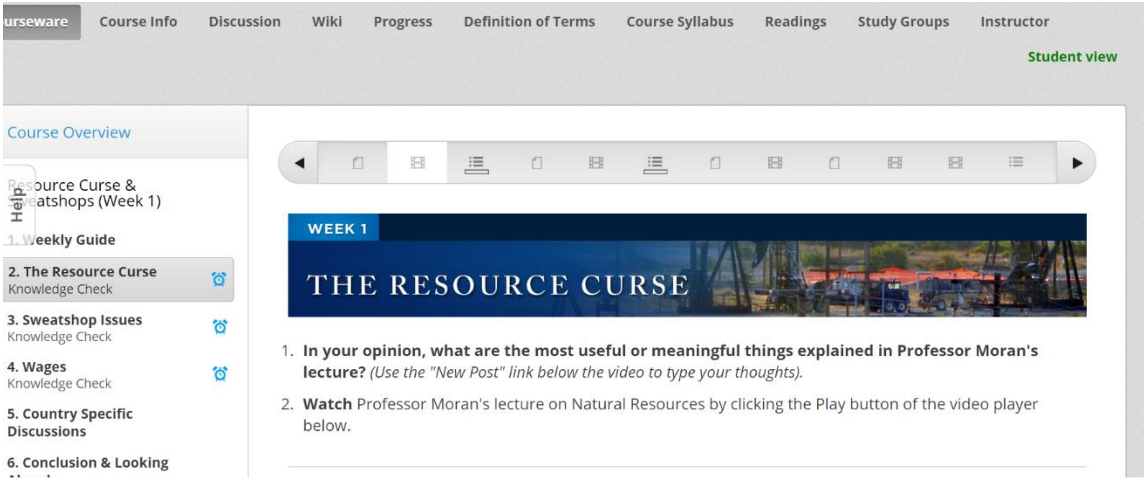

Figure 2: Cognitive Presence Sequence

After the triggering event (questioning prompt show in Figure 1), learners were able to interact with the video and take notes. The video itself included key term prompts as pop-up bubbles which functioned as another triggering event within the video components (Figure 2).

Figure 3: Key Term within video component

Also included in Table 1 are the design elements and indicators used for the integration and resolution phases of cognitive presence. After learners engage with the video and take notes, they are asked to participate in self-assessments, engage in peer-to-peer discussions, and complete knowledge checks. Achievement is measured based on the scores they receive from the self-assessment and knowledge checks. The learning sequences follow a similar pattern over the duration of the seven-week course. This learning sequence consistency in design enabled us to examine learner forum posts at the overall course level in relation to key term use, language abstraction, and video activity enabling us to better understand the exploration phase of the learning sequences, specifically, to better understand the relationship between exploration and achievement.

3. Methods

The data sets used in this case were from the pre-survey completed by participants who registered in the GeorgetownX *Globalization* MOOC offered in 2013 along with activity data from MOOC participation. After cleaning the pre-survey dataset to remove participants under the age of 18, responses with missing data for key variables, and respondents with less than full professional English language proficiency, our final dataset comprised 495 learners. We then extracted the course data from edX concentrating on the variables described in Table 1 specifically for examining exploration, which included:

- Number of discussion forum posts made in the course
- Average Length of Discussion Posts (words count)
- Overall video activity in course (Video activity was obtained by summing the number of video-related events: play, pause, seek, change playback speed, recorded for each student.)
- Overall course grade/score

In addition to the data listed above, we also wanted to determine whether language abstraction and use of key terms in discussion forum posts related to the course grade/score. Miaomiao, Yang, and Rosé (2014) measured the level of cognitive engagement in their study of motivation and cognitive engagement in MOOCs by calculating a numerical rating of abstractness of a word using the publicly available Abstractness dictionary (Turney et al., 2011) and computed the mean level of abstraction for each post by adding the abstractness score of each word in the post and dividing it by the total number of words. This was undertaken working on the assumption and precedents in the literature, that level of language abstraction reflects the understanding that goes into using those abstract words when creating the post, and thus shows a higher level of cognitive engagement.

We expanded on Miaomiao, Yang, and Rosé's methodology by also including the use of key terms derived from core concepts addressed in the course, which we identified with the content experts as part of the instructional design process. By analyzing the discussion forum posts in relation to the key terms and examining the level of activity of learners in relation to the video components of the course we aim to understand learners' exploration of the course content in relation to achievement.

3.1 Input variables potentially related to score were gathered

The analysis was performed through the statistical programming language R, using linear regression and ANOVA analysis.

To examine the relationship between key term use and language abstraction on student achievement, we accounted for other factors that could affect student achievement. To that end, model selection procedures were used to test a wide array of input factors and identify those that explain the bulk of the variation among student scores. The resulting model would provide the most reliable picture of the relationship between key term use, language abstraction, and student achievement.

With this in mind the following steps were taken:

1. input variables potentially related to score were gathered,
2. linear regression models for student achievement based on combinations of input variables were created and tested,

3. an ultimately best model was selected, and
4. inference regarding the relationships between input variables and student score was performed.

These steps are described in more detail below.

In order to account for other factors potentially related to student achievement, data from pre-course surveys was combined with course activity, discussion post and overall achievement (score) data. The following factors were included in the analysis:

Variables for Analysis

- Student Achievement
- Key Term Use Score
- Abstraction Language Score
- Video Activity
- Number of Discussion Posts Made
- Average Length of Discussion Posts (words)
- Overall Course Activity - related to how many of the chapters/sections the student was active in the course based on Navigation, Video, or Problem clicks (events in the edx log). If the student was active in six or more chapters out of nine total, the Activity Threshold Variable was a 1, if they were active in less than six chapters, the Activity Threshold Variable was a 0

Self-reported factors from the pre-course survey included:

- Interest in Topic
- Interest in Learning Objective
- Intrinsic Motivation (quantified based on selection to specific questions in the survey)
- Extrinsic Motivation (quantified based on selection to specific questions in the survey)
- Importance of Receiving a Certificate for the Course
- Technological Aptitude
- English Level
- Age
- Education
- Employment

3.2 Linear regression models for student achievement based on combinations of input variables were created and tested

To isolate the effects of key term use and language abstraction on student performance, each of the above variables was included in the models. In order to find the optimal combination of the above variables and the interaction terms, variables and interaction terms were tested with backward-forward stepwise regression. Backward-forward stepwise regression walks through subsets of all input variables, based on removing, testing and adding variables as a function of their statistical significance, starting with the full set of input variables. It does this by considering all input variables, determining which variable is least significant, removing it, then re-testing and again removing the least significant variable. At this point, it considers returning the previously removed variable to the model, in case the exclusion of one variable has made a previously excluded variable significant again. The removal or addition of a variable is based on statistical significance metrics. This method works well when many variable combinations need to be tested. In cases where the input variable was an abstract quantity (key term score or intrinsic motivation, for example), the natural logarithm of the variable was used to aid in interpretation; meaning that we examined the expected change in student performance based on a percentage change in the variable. For clearly interpretable input variables like age, or education level, input variables were not transformed.

The 'best model' was determined by the Akaike Information Criterion and the Adjusted R-Squared value – with the goal of maximizing model accuracy while including a 'penalty' for too many input variables to reduce overfitting and allow for model interpretation. Lasso regression was also performed based on all combinations of variables and important interaction terms to identify key variables to compare with the results of the exhaustive stepwise regression. Qualitative variables with many levels were examined using individual variable ANOVA analysis and testing within the model in order to determine the optimal number of factor levels.

3.3 A best model was selected

After analysis of the results, a best model was selected with the goal of isolating the effects of key term use and language abstraction on student achievement, while simultaneously including as many other significant input variables as possible while maintaining interpretability.

3.4 Inference regarding the relationships between input variables and student score was performed

Once the best model was selected based on the above criteria, this model was used to examine relationships between key term use, language abstraction and other input variables on student achievement.

4. Results

Based on the best model selected, as described in the Methods section, we found several factors that were statistically significant in relation to student achievement. The statistically significant variables relating to linguistic analysis for language abstraction and key terms use and video activity analysis are presented in this section. The summary statistics of the full best model are also presented here; the full best model, including all coefficients and p-values are presented in the Appendix.

As our analysis is exploratory, we considered a range of significance levels ranging from 99.9% to 90%, in combination with the coefficient values found, in order to examine the relationships between student achievement, key term use, language abstraction, and video activity. Given the variance in student learning, we used the sign and relative size of the coefficients, in combination with their statistical significance, to determine which input variables positively and negatively affect student achievement, as well as the relative magnitude of the effects. Where we had moderately high statistical significance levels, we used the coefficients to gauge and rank the effect of the input variables on student achievement. Where we had high statistical significance levels, we used the coefficients to estimate the change in expected achievement score as a function of a change in the input variable. In general, we found moderate statistical significance levels regarding linguistic data (language abstraction score and key term use), and high statistical significance levels regarding the video activity data.

The overall best model fits the student achievement data well. Specifically, incorporating the variables Intrinsic Motivation, Extrinsic Motivation, Self Expectation of Achievement, Technological Aptitude, Age, Education, Employment, Abstraction Score, Key Word Score, Average Length of Posts, Number of Posts Made, Activity Level (active in 80% of the course or more, Y/N), Video Activity, and the appropriate interaction terms, 91.4% of the variance in student performance is explained. With an F-statistic of 204 and corresponding p-value of < 2.2e-16, the overall best model provides information about student performance at above the 99.9% confidence level. Table 2 shows the overall best model summary statistics.

Overall Best Model Summary Statistics
Multiple R-squared: 0.9206, Adjusted R-squared: 0.916
F-statistic: 200.6 on 27 and 467 DF, p-value: < 2.2e-16

Table 2: Overall Best Model Summary

Most of the variance in student achievement is based on course activity, but the addition of the linguistic data, video activity, and other factors as presented in the best model (Table 3) explain additional variance in score at above the 99.9% confidence level.

Model 1: Student Achievement ~ Course Activity Level						
Model 2: Student Achievement ~ Overall Best Model						
	Res.Df	RSS Df	Sum of Sq	F-statistic	P-value	Significance
Model 1:	493	8.14				
Model 2:	468	6.25	1.89	5.67	1.37e-15	***

Table 3: Model comparison using Analysis of Variance (ANOVA) test

4.1 Key Term Use and Language Abstraction

Linguistic data based on student discussion posts were found to be statistically significant. Also, a number of interactions are statistically significant, as there is the potential for a lot of interplay between the length of a post and the types of words used by a student. The variables analyzed should be interpreted in terms of percent change - i.e. "a 10% percent change in Abstraction score could potentially lead to a coefficient/10 point change in expected student performance." Table 4 summarizes the combined effects of the individual input variables.

Input Variable	Coefficient	P Value	Significance
Language Abstraction Score	-2.09	0.02	*
Average Length of Posts	0.29	0.01	**
Key Term Use Score	45.24	0.02	*
Language Abstraction Score X Average Length of Posts	0.51	0.02	*
Language Abstraction Score X Key Term Score	75.98	0.03	*
Average Length of Posts X Key Term Score	-10.31	0.03	*
Number of Posts Made X Key Term Score	-1.64	0.10	.
Language Abstraction Score X Average Length of Posts X	-17.46	0.03	*

Key Term Score			
Average Length of Posts X Number of Posts Made X Key Term Score	0.36	0.10	.

Table 4: Combined effects of input variables
Significance codes: 99.9 % significant '*' ; 99 % significant '**' ;95 % significant '*' 90% significant '.'**

Combining the effects of each input variable highlights the importance of key term use and language abstraction score as shown in Table 5.

Total expected effect on student achievement of linguistic variables	Combined Effect
Language Abstraction Score	56.94
Key Term Score	92.17
Average Length of Posts	-26.61
Number of Posts Made	-1.28

Table 5: Linguistic variables and student achievement
Statistically Significant at the 90% or above level

4.2 Video Activity and Student Achievement

Video activity is a statistically significant factor relating to student achievement. Based on the analysis presented, holding Intrinsic Motivation, Extrinsic Motivation, Self Expectation of Achievement, Technological Aptitude, Age, Education, Employment, Abstraction Score, Key Word Score, Average Length of Posts, Number of Posts Made and Activity Level constant - an increase in video activity correlates with increase in student performance.

Input Variable	Coefficient	P Value	Significance
Overall Course Activity	0.850	2e-16	***
Overall Video Activity	0.046	2e-16	***
Overall Course Activity X Overall Video Activity	-0.035	0.002	**

Table 6: Activity and Student Achievement
Significance codes: 99.9 % significant '*' ; 99 % significant '**' ;95 % significant '*' 90% significant '.'**

The reason that there is a negative coefficient for the variable Overall Course Activity X Overall Video Activity is because there is slight 'overlap' between overall activity and video activity; participants who are more active in the course also have higher video activity levels. It is important to include the interaction to account for the relationship between the variables and explore further.

Given the statistical significance levels shown in Table 6, we proceeded to examine the change in expected student achievement based on change in video activity for High Overall Course Activity students and for Low Overall Course Activity students. Because the course activity variable used was an indicator course activity threshold variable - we show in Table 7 the different effects of video activity on students in the high and low activity groups.

For students who were active in 80% or more of the course (high activity), a 10% increase in video activity could potentially lead to a 1.1% increase in score, whereas for students in the lower course activity group, a 10% increase in video activity could potentially lead to a 4.6% increase in score. Table 7 shows the results of this analysis, combined effects of the coefficients shown in Table 6, and Appendix 1 includes the Overall Best Model for further detail.

Total expected effect on student achievement of video activity and course activity	Combined Effect
Video Activity for High Course Activity Students	0.011
Video Activity for Low Course Activity Students	0.046

Table 7: Effect on student achievement of video activity and overall course activity

4.3 Video Activity and Linguistic Data Correlation

In order to investigate the relationship between video activity and linguistic data, the Pearson correlation coefficient between Video Activity and Language Abstraction Score, Key Term Use Score, Number of Posts Made, Average Length of Post was calculated, holding all else constant. Specifically, video activity was regressed on all other factors except for Language Abstraction Score, Key Term Use Score, Number of Posts Made and Average Length of Posts; the best model for Video Activity was examined using exhaustive stepwise regression as described above, and the residuals were used to calculate the Pearson correlation coefficient. Table 8 shows the results of this analysis. Key Term Use Score is most correlated with video engagement, followed by the Number of Posts Made.

Video Activity and Linguistic Data Correlation	Pearson Correlation	P-value	Significance
Language Abstraction Score	-0.01	0.82	
Key Term Use Score	0.28	1.26e-10	***
Number of Posts Made	0.19	3.41e-05	***
Average Length of Posts	0.12	0.01	**

Table 8: Video Activity and Linguistic Data Correlation
Significance codes: 99.9 % significant '*' ; 99 % significant '**' ;95 % significant '*' 90% significant '.' Not significant [blank]**

The relevance of linguistic data to video activity is evident when comparing the best model for video activity using everything except for linguistic data versus the best model for video engagement including linguistic data.

The best model for video activity excluding linguistic information, created through the methods described above, included Overall Course Activity and Interest in Learning Objective - the other input variables (described earlier in this section) were not statistically significant. With an Adjusted R Squared of .336, this model explains 33.6% of the variance in video engagement.

The best model for video activity including linguistic information explains 42.8% of the variance in video engagement with an Adjusted R Squared of .428. The model including linguistic information explains more variance in video engagement at the 99.9% confidence level. Table 9 shows the results from the comparison of these two models.

Model 1: Video Activity ~ Best Model without Linguistic Data							
Model 2: Video Activity ~ Overall Best Model including linguistic data							
	Res.Df	RSS	Df	Sum of Sq	F-statistic	P-value	Significance
Model 1:	492	80974242					
Model 2:	477	67644811	15	13329431	6.2662	3.68e-12	***

Table 9: Video Activity models comparison

The analysis and results in this section offers insights into the relationships between student achievement, key term use of core concepts in the course, language abstraction, and video activity.

However, further analysis would be helpful to go beyond this exploration. It appears that there may be two distinct populations within the data - based on score and activity levels. Examining these two groups individually could yield more detailed results and further understanding of participants' cognitive presence in the course.

5. Conclusions

This case study examined aspects of participants' cognitive presence in a GeorgetownX MOOC on globalization. The impetus for this case study was to gain a better understanding of how MOOC participants engage with the course materials by analyzing both linguistic and activity variables. We believe that this combination will enable us to start mapping the depth of learner cognitive presence within a MOOC and will in the future enable us to make better learning design decisions. With this in mind, in this section, we highlight some of the key findings stemming from the analysis included in the Results section of this paper.

One of the most interesting results is the large effect of Language Abstraction Score and Key Term Use Score. As we examine the coefficients for statistical significance, sign, and relative size, we can see that Key Term Use Score and Language Abstraction Score are related to student achievement. It seems that there is a positive relationship between higher achievement in the course and students' usage of more key terms relating to core concepts and high language abstractness in their writing. Holding all other variables described in the Methods and Results sections constant, an increase in the number of key terms used by a student is related to an increase student achievement. An increase in a student's abstraction levels in their discussion posts is also positively correlated to an increase in score, although to a lesser degree. Given that the assessment design elements used in the course were directly aligned with the course core concepts addressed in the videos, this result supports our course design decision to enable exploration of video content by having triggering events (questioning prompt and key term video bubble) that serve to encourage students to take notes using the discussion forum posts. However, writing more (average length of posts) doesn't show the same relationship with student achievement.

Surprisingly, we saw a negative coefficient on Number of Posts Made (Table 5) in relation to student achievement given that, in general, higher participant activity in the course typically accompanies better student achievement. This result could potentially be explained because the activity level in the course is such a major factor regarding expected student achievement that it has already accounted for the increase in expected student achievement related to higher participation. In terms of video activity, based on the analysis shown in Table 7, a 10% increase in Video Activity is correlated with an increase in student performance of 1.1 points for students who had high activity in the course and 4.6 points for students with low activity in the course, on a 0 to 100 point scale. For example, we would expect that a 10% increase in video activity, would be accompanied by an increase in score from 74.5 to 75.6 out of 100.

Note that in this course the passing grade was set at 75%, so finding ways as part of the design of the learning sequence to trigger further video interaction could be of benefit to those participants who are close to the passing grade. For this to happen, however, both participants and instructors would need access to a monitoring system within MOOC platforms that makes visible the learning sequence itself in relation to critical factors that influence achievement. This type of monitoring system could then enable targeted interventions to take place and learning analytics would in and of itself become part of the learning sequence design to encourage depth of participants' cognitive presence.

6. References

Boris, G., & Hall, T. (2005). *The first decade of the community of inquiry framework: A retrospective. 20th Annual Conference on Distance Teaching and Learning*. http://www.uwex.edu/disted/conference/Resource_library/proceedings/04_1288.pdf

Garrison, D. R, Anderson, T., and Archer, W. (2010). The first decade of the community of inquiry framework: A retrospective. *Internet and Higher Education*, 13:5-9.

Koller, D., Ng, A., Do, C., & Chen, Z. (2013). Retention and intention in massive open online courses: In depth. *Educause Review*, 48(3), 62–63.

Sheu, F. R., Lee, M. M., Bonk, C. J., & Kou, X. (2013, June). *A mixed methods look at self-directed online learning: MOOCs, open education, and beyond*. Paper presented at the 25th Annual Ethnographic & Qualitative Research Conference. Cedarville, OH.

Turney, P.D., Neuman, Y., Assaf, D., and Cohen, Y. (2011). Literal and metaphorical sense identification through concrete and abstract context, *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP-2011)*, Edinburgh, Scotland, UK, July 2011, pp. 680-690.

Wen, M., Yang, D., and Rose, C. P. (2014). Linguistic Reflections of Student Engagement in Massive Open Online Courses. *ICWSM'14*, 2014.

Appendix 1: Overall Best Model

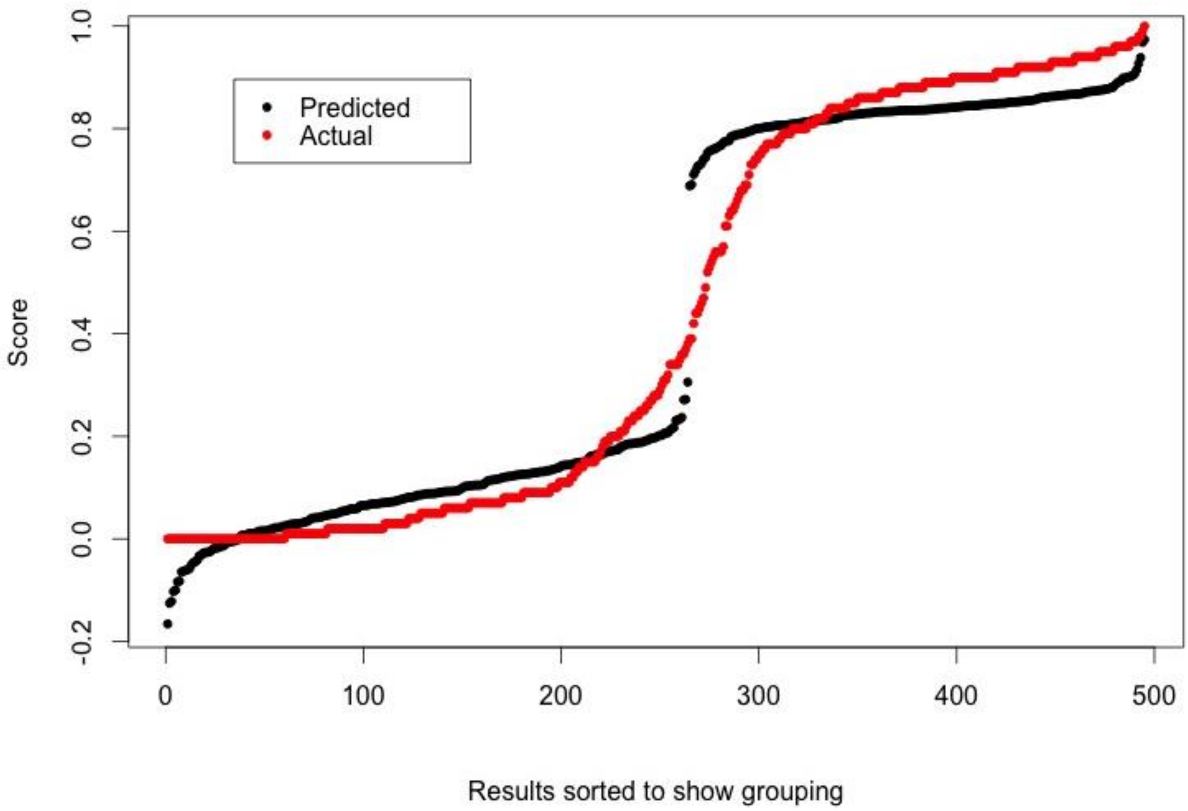

Figure 4: Predicted (black) and actual (red) scores plotted against rank to allow comparison.

Input Variable	Coefficient	P-Value	Significance
(Intercept)	-2.36	0.00	***
Intrinsic Motivation	-0.04	0.35	
Extrinsic Motivation	-0.09	0.06	.
Self Expectation of Achievement: Complete Course and Receive Certificate	0.11	0.00	**
Self Expectation of Achievement: Complete Course and Not Receive Certificate	0.13	0.00	**
Self Expectation of Achievement: Participate only Chapters I'm Interested In	0.06	0.21	
Technological Aptitude	0.32	0.03	*
Age	0.02	0.04	*
Education Level: High School or Below	1.20	0.08	.
Employment: Retired	0.03	0.19	
Employment: Homemaker	-0.04	0.17	
Abstraction Score	-2.09	0.02	*
Average Length of Posts	0.29	0.01	**
Number of Posts Made	0.03	0.69	
Key Word Score	45.24	0.02	*
Activity Threshold	0.85	0.00	***
Video Activity	0.05	0.00	***
Activity Threshold X	-0.04	0.00	**

Video Activity			
Intrinsic Motivation X Extrinsic Motivation	0.11	0.09	.
Technological Aptitude X Age	-0.01	0.03	*
Technological Aptitude X Education Level: High School or Below	-0.38	0.08	.
Abstraction Score X Average Length of Posts	0.51	0.02	*
Average Length of Posts X Number of Posts Made	-0.01	0.77	
Abstraction Score X Key Word Score	75.98	0.03	*
Average Length of Posts X Key Word Score	-10.31	0.03	*
Number of Posts Made X Key Word Score	-1.64	0.10	.
Abstraction Score X Average Length of Posts X Key Word Score	-17.46	0.03	*
log(average.len.post+1):Number of Posts Made X Key Word Score	0.36	0.10	.

Table 10: Overall Best Model, Student Performance, Coefficients and Significance

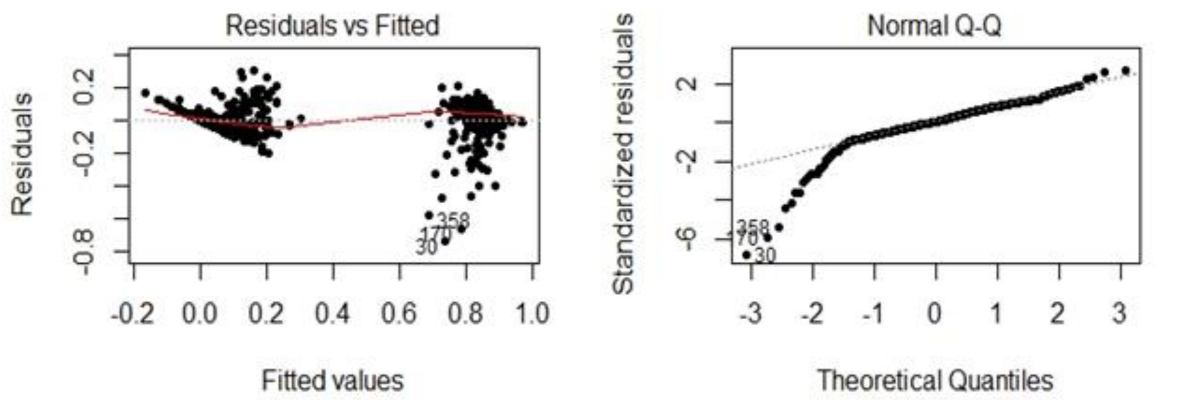

Figure 5: Residuals vs. Fitted Values & Normal Q-Q Plot

About this Paper

Acknowledgements

This document was produced with funding from the European Commission Seventh Framework Programme as part of the LACE Project, grant number 619424.

Citation details

Vovides, Y., Youmans, T., Arthur, P., Davis, D., Ayo, E., Pongsajapan, R., McWilliams, M. and Kruse, A. LAK15 Case Study 2: Examining Learners' Cognitive Presence in Massive Open Online Courses, *Learning Analytics Review*, Paper LAK15-2, ISSN 2057-7494, March 2015, http://www.laceproject.eu/learning-analytics-review/examining-learners-cognitive-presence-in-moocs/

For more information, see the LACE Publication Policy: http://www.laceproject.eu/publication-policy/. Note, in particular, that some images used in LACE publications may not be freely re-used.

Please cite this document including the issue number (LAK15-2) and the Learning Analytic Review's ISSN (2057-7494). The Learning Analytics Review is published by the LACE project at the University of Bolton, Bolton, UK.

The persistent URL for this document is: http://www.laceproject.eu/learning-analytics-review/examining-learners-cognitive-presence-in-moocs

About the Authors

Yianna Vovides is Director of Learning Design and Research at Georgetown University's Center for New Designs in Learning and Scholarship (CNDLS). She is also faculty in the Communication, Culture, and Technology Program, Graduate School of Arts and Sciences at Georgetown University. Yianna has over 15 years of experience in instructional design and technology. Her experience comes from both academic and professional practice. Her research focus is on the use of learning analytics to support teaching and learning within cyberlearning environments.

Thomas Youmans is a graduate associate at the Center for New Designs in Learning and Scholarship (CNDLS) at Georgetown University. He is master's student in mathematics and statistics, and uses mathematical methods and statistical principles to model and investigate how complex systems function and interact. Tom's research focuses on the examination of educational and learning models based on measurable information about learner characteristics and behaviour in the online context.

Paige Arthur is a graduate associate at the Center for New Designs in Learning and Scholarship (CNDLS) at Georgetown University. A master's student in language and communication, Paige uses discourse analysis and other linguistic methodologies to research the way that people tell stories and reflect online and through talk. Current research examines student writing and reflection in online courses and assignments, and how university alumni make sense of their personal development and professional pathways to construct self-narratives.

Daniel Davis is a graduate associate at the Center for New Designs in Learning and Scholarship (CNDLS) at Georgetown University. Dan is in his second year in the Communication, Culture & Technology Master's program with a focus on Learning Sciences and Technology Design. His interest in education was sparked by his time spent at the National Education Association, where worked on education policy and advocacy. At CNDLS, Dan works with the Learning Design and Research team in exploring learning analytics and instructional design technologies.

Rob Pongsajapan is Assistant Director for Web Projects at Georgetown University's Center for New Designs in Learning and Scholarship, where he works on communication projects with data from Georgetown's MOOCs. He also works closely with faculty and students on ePortfolio projects, and is interested in the ways people engage with and reflect on the data they produce.

Mindy McWilliams is Associate Director for Assessment at Georgetown University's Center for New Designs in Learning and Scholarship, where she assesses the impact of teaching with technology, curricular impact, and most recently, non-cognitive aspects of learning. She is co-PI on Georgetown's Formation by Design Project, which promotes a learner-centered, design-based approach to reinventing our institutions around whole person development. The Project engages internal and external stakeholders in a process of defining, designing, and measuring formation of the individual within the context of higher education.

Anna Kruse is Assistant Director for Strategic Integration and Communication at Georgetown University's Center for New Designs in Learning and Scholarship (CNDLS). She has played many roles at CNDLS, most recently managing both the Initiative on Technology-Enhanced Learning (ITEL), a University-wide initiative to encourage creativity and experimentation around the use of technology for teaching and learning, and Georgetown's partnership with edX to develop MOOCs

Licence

About the Learning Analytics Review

Background

The Learning Analytics Review provides a series of stand-alone series of articles aimed primarily at people who want to make decisions about what they are going to use learning analytics. While they will be of an authoritative and scholarly character, they will generally be white papers or briefings. The white papers and briefings are complemented by additional papers related to various aspects of learning analytics which will be of interest to the broad learning analytics community.

About this Learning Analytics Review Paper

To support the LACE project's community-building work a series of three papers have been published based on sessions which were presented at the LAK 15 conference. These are:

1. Kuzilek, J., Hlosta, M., Herrmannova, D., Zdrahal, Z., and Wolff, A. LAK15 Case Study 1: OU Analyse: Analysing at-risk students at The Open University, *Learning Analytics Review*, Paper LAK15-1, ISSN 2057-7494, March 2015, http://www.laceproject.eu/learning-analytics-review/analysing-at-risk-students-at-open-university/
   This paper was scheduled to be presented on 18th March 2015 in the Students At Risk session and on 19th March 2015 in the Technology Showcase session.
2. Vovides, Y., Youmans, T., Arthur, P., Davis, D., Ayo, E., Pongsajapan, R., McWilliams, M. and Kruse, A. LAK15 Case Study 2: Examining Learners' Cognitive Presence in Massive Open Online Courses, *Learning Analytics Review*, Paper LAK15-2, ISSN 2057-7494, March 2015, http://www.laceproject.eu/learning-analytics-review/examining-learners-cognitive-presence-in-moocs/
   This paper was scheduled to be presented on 18th March 2015 in the MOOCs—Discussion Forums (Practitioner) session.
3. Grann, J. LAK15 Case Study 3: Flexpath: Building Competency-based, Direct Assessment Offerings, *Learning Analytics Review*, Paper LAK15-3, ISSN 2057-7494, March 2015, http://www.laceproject.eu/learning-analytics-review/building-competency-based-offerings/
   This paper was scheduled to be presented on 19th March 2015 in the Learning Strategies and Tools session.

About the LACE project

The LACE project brings together existing key European players in the field of learning analytics & educational data mining who are committed to build communities of practice and share emerging best practice in order to make progress towards four objectives.

> *Objective 1 – Promote knowledge creation and exchange*
> *Objective 2 – Increase the evidence base*
> *Objective 3 – Contribute to the definition of future directions*
> *Objective 4 – Build consensus on interoperability and data sharing*

For more information, see the LACE web site at http://www.laceproject.eu/

http://www.laceproject.eu *@laceproject*