2019

Setting The Pace: Examining Cognitive Processing in MOOC Discussion Forums With Automatic Text Analysis

Robert L. Moore
Old Dominion University

Kevin M. Oliver

Chuang Wang

Original Publication Citation

Moore, R. L., Oliver, K. M., & Wang, C. (2019). Setting the pace: Examining cognitive processing in MOOC discussion forums with automatic text analysis. *Interactive Learning Environments, 27*(5-6), 655-669. doi:10.1080/10494820.2019.1610453

Setting the Pace: Examining Cognitive Processing in MOOC Discussion Forums with Automatic Text Analysis

Robert L. Moore[a]*, Kevin M. Oliver[b] , and Chuang Wang[a]

[a]Department of STEM Education & Professional Studies, Old Dominion University, Norfolk, USA; [b]Teacher Education and Learning Sciences, North Carolina State University, Raleigh, USA; [a]Department of Educational Leadership, University of North Carolina Charlotte, Charlotte, USA

*https://orcid.org/0000-0002-5645-9297. http://www.twitter.com/robmoore3

Robert L. Moore is an Assistant Professor of Instructional Design & Technology at Old Dominion University. His research interests include the design and development of online learning environments and the role of analytics in exploring student engagement in these environments.

Kevin M. Oliver is a Professor of Learning Design and Technology at North Carolina State University. His research interests include online/distance learning pedagogy with implications for instructional design/policy.

Chuang Wang is a Professor of Educational Research Measurement and Evaluation at the University of North Carolina Charlotte. His research interests include multi-level modelling and educational statistics.

Setting the Pace: Examining Cognitive Processing in MOOC Discussion Forums with Automatic Text Analysis

Learning analytics focuses on extracting meaning from large amounts of data. One of the largest datasets in education comes from Massive Open Online Courses (MOOCs) that typically feature enrollments in the tens of thousands. Analyzing MOOC discussion forums presents logistical issues, resulting chiefly from the size of the dataset, which can create challenges for understanding and adequately describing student behaviors. Utilizing automatic text analysis, this study built a hierarchical linear model that examines the influence of the pacing condition of a massive open online course (MOOC), whether it is self-paced or instructor-paced, on the demonstration of cognitive processing in a HarvardX MOOC. The analysis of 2,423 discussion posts generated by 671 students revealed the number of dictionary words used were positively associated with cognitive processing while analytical thinking and clout was negatively associated. We found that none of the student background information (gender, education), status of the course engagement (explored or completed), or the course pace (self-paced versus instructor paced) significantly influenced the cognitive processing of the postings.

Keywords: MOOCs; automated text analysis; learning analytics; discussion forums; community of inquiry; cognitive presence; LIWC

1. Introduction

There is no all-encompassing definition of a Massive Open Online Course, or MOOC, in the literature. Instead, the term is used to refer to a variety of offerings (Major & Blackmon, 2016). And no consensus exists regarding the number of types of MOOCs, ranging from two (e.g. Major & Blackmon, 2016) to three types (Bonk, Lee, Reeves, & Reynolds, 2017) and others suggesting that only two is too limiting (Cohen & Holstein, 2018; Lowenthal, Snelson, & Perkins, 2018). Two of the common types are cMOOC and xMOOC. A key distinction between these two types is the role of the learner in the learning process (Hew & Cheung, 2014; Major & Blackmon, 2016). In a cMOOC or connectivist MOOC, the expectation is for the learner to not only participate in discussions, but to take an active role in the production, discovery, and debate about course content. A distinguishing characteristic of a cMOOC is the use of various external tools, such as Twitter or blogs, produced by students outside of the course site but brought into the course space by the students (Bonk et al., 2017). In a cMOOC, learning is influenced by openness, autonomy, and connections across a network of distributed knowledge (Major & Blackmon, 2016; Siemens, 2005). In contrast, xMOOCs are more teacher-driven, and instruction is primarily in the form of pre-recorded video lectures, self-graded assessments and discussion forum activities (Adams, Yin, Vargas Madriz, & Mullen, 2014; Bonk et al., 2017) and use a behaviorist view of learning that Hayes (2015) defines as a knowledge transmission model. While the initial MOOCs were mostly classified as cMOOCs, more xMOOCs are now being offered, primarily by traditional higher education institutions (Al-Imarah & Shields, 2018; Veletsianos & Shepherdson, 2016).

Currently, there are two pacing conditions among MOOCs: instructor-paced and self-paced. An instructor-paced MOOC, also referred to as a live, session-based, or cohort MOOC, will have a specific start and end date and a defined enrollment period. These courses will have a release date (when the course is opened to students) and specific deadlines, like one would find in an on-campus course. While students may join at any time, many will start on the launch date and form a cohort. This approach has been linked to higher student retention rates (Sharif & Magrill, 2015). When an instructor-paced course has passed the final deadline, it is archived. Learners can still sign up for the course, but they will not earn a certificate. Alternatively, there are self-paced or on-demand MOOCs. These courses have everything released at once and there are no deadlines. Most self-paced courses do still end at some point, for practical reasons (e.g. forum support and research cycles), and these can also be archived when they end. The major providers of MOOCs, edX and Coursera, both offer MOOCs in the self-paced structure. Typically, these self-paced MOOCs feature the same content and activity requirements as their instructor-paced counterparts, with the most significant difference being the loss of the cohort moving through the course at the same time. Additionally, many of the self-paced MOOCs do not offer certificates of completion, though some do feature a teaching assistant with a presence in the discussion forums. In this study, the researchers identified a course offered in both instructor-paced and self-paced conditions, allowing an opportunity to make comparisons between the two. Data from these courses were provided through a data use agreement between the researcher's home institution and Harvard University's VPAL Research Team.

2. Theoretical background

The Community of Inquiry (CoI) theoretical framework informs this research (Garrison, Anderson, & Archer, 2001). This framework, comprised of domains for teaching, social and cognitive presence, suggests high-quality online education is realized when learners interact socially and coordinate efforts with peers, link new knowledge to past understanding, and directly apply the information to their present lives in a learning environment that places great importance on self-reflection and self-regulated learning (Garrison, 2007; Kilis & Yıldırım, 2018; Shea & Bidjerano, 2010). This framework has become the most widely cited model for understanding learning through online discussions (Breivik, 2016) and is viewed as an "essential context for higher-order learning" (Garrison et al., 2001, p. 7). In the MOOC context, the community of inquiry primarily operates in the course discussion forum. CoI incorporates a constructivist approach to learning and recognizes that interaction and discourse play a fundamental role in higher-order learning, along with structure (design) and leadership (facilitation and direction) (Garrison, 2007).

2.1. Discussion forums

Extant MOOC research tends to emphasize forum activity, as it is the primary source of student engagement. Within forums, students are afforded space for peer connection and idea exchange and can engage in conversation and interact with one another (Beckmann & Weber, 2016; Kent, Laslo, & Rafaeli, 2016; Sharif & Magrill, 2015). Asynchronous forums allow time for reflection and the creation of a written record, constituting an effective mechanism for students to build on each other's ideas (McLoughlin & Mynard, 2009; Wang, Woo, & Zhao, 2009).

One should look at cognitive presence as a process that fosters the development of higher-order thinking skills instead of individual learning outcomes (Akyol et al., 2009; Akyol & Garrison, 2011). In forums, students can collaborate in knowledge construction that leads to higher levels of cognitive presence (Kent et al., 2016). McLoughlin and Mynard (2009) found that several elements of discussion forums supported the development of higher-order thinking, such as time to reflect on one's writing and the creation of a written record that allows students to more effectively build upon each other's ideas. These researchers further suggest that the type of assigned task and the wording of the initial discussion forum prompt can affect the type of higher-order thinking processes that will emerge in online discussion. Forums alone will not likely support greater cognitive presence but can be optimized to do so. Garrison (2007) explains that cognitive presence is operationalized through Dewey's (1933) practical inquiry model on reflective thinking and has four phases: triggering event, exploration, integration, and resolution. To the extent a forum is structured to encourage these phases, it is more likely to support cognitive presence.

3. Learning analytics

Online discussion forums, particularly in MOOCs, are the spaces where a significant amount of student-student interaction takes place (Wong, Pursel, Divinsky, & Jansen, 2015). The different approaches and frameworks that have been developed to measure critical thinking have been met with various levels of support or criticism, with the Community of Inquiry (CoI) framework emerging as one of the more common (Breivik, 2016). While the CoI model is well-researched and validated, it can be tedious to hand-code responses under this approach (Kovanović et al., 2016). This is a feasible approach when the dataset for a study is smaller or when there is a team of researchers, but it is far less feasible for studies looking at MOOC discussion forums, given the volume of posts to be analyzed. In response, researchers have examined MOOC discussion forums to look for evidence of cognitive presence using automatic text analysis tools (Dowell & Graesser, 2014; Kovanović et al., 2016).

Learning analytics, specifically text mining, can be used in both unsupervised and supervised machine learning approaches to make analysis more manageable

(Moore, 2019). Ezen-Can, Boyer, Kellogg, and Booth (2015) observed a marked increase in interest in using learning analytics to better understand student activities within MOOCs. Specifically, the researchers state that "one very important source of data in MOOCS is the textual dialogue among students . . . on discussion forums" (p. 146).

In this study, our interest was in discussion posts, which Humphreys and Wang (2018) believe are ideal candidates for using automated text analysis because it makes comparison between groups possible while handling an expansive corpus. The ability to leverage powerful and efficient tools for analysis has led many researchers to deploy automated text analysis tools to examine online interactions (Donohue, Liang, & Druckman, 2014).We selected the extensively validated Linguistic Inquiry Word Count (LIWC) tool for analyzing our dataset (Fast, Chen, & Bernstein, 2017; Tausczik & Pennebaker, 2010) (pronounced “Luke”). This validated tool uses dictionaries to categorize and quantify language used in text and provides a calculation of the percentage of words within defined categories (Khazaei, Lu, & Mercer, 2017; Simms et al., 2017).

We selected the 2015 version of the LIWC tool to analyze our discussion forum text. As Kahn et al. (2007) explain, LIWC compares text to pre-determined dictionaries with specified categories and outputs a numerical frequency value of post content along these categories, ranging from simple articles and prepositions to nuanced emotion, affective, and cognitive words. Pennebaker, Boyd, Jordan, and Blackburn (2015) use the term “target words”, which they define as the words analyzed by LIWC. In this study, the target words are those found in the discussion forum posts. The words found in LIWC dictionaries are called “dictionary words”, and the term “word categories” refers to groups of words that describe a specific domain, such as “insight” or “discrepancy” (Pennebaker et al., 2015). For this study, we focused on cognitive presence which we are operationalizing using the cognitive processing category in LIWC. This category is made up of a total of 797 words that can be further divided into sub-categories of “insight”, “cause”, “discrepancies”, “tentativeness”, “certainty”, and “differentiation” (Table 1). The words that appear in these subcategories (e.g., “insight”, “causation”) are also in the main word category (“cognitive processes”). LIWC’s usefulness has been validated through various studies, including those looking at cognitive presence (Kovanović et al., 2016; Oztok, Zingaro, Brett, & Hewitt, 2013; Pennebaker et al., 2015), and its cognitive processing score has been found to have high levels of predictive validity (Slotter & Ward, 2015; Tausczik & Pennebaker, 2010). LIWC yields scores on four summary variables and also indicates what percentage a particular language category accounts for in relation to total number of words (Krieger, Watkins, Gerber, Pham, & Bauman, 2018). LIWC then reports a value, expressed as a percentage, for those words. For example, a score of 8.3 for a post’s cognitive processing means that 8.3 percent of the words used in the post were in the cognitive processing dictionary (Simms et al., 2017). This use of a top-down dictionary analysis approach allows for more consistent measurement of words, and the use of a validated tool allows for concurrent validity (Humphreys & Wang, 2018).

Table 1. LIWC Word Categories

Category	Examples	# of Words in Category
Cognitive processes	Since, seem, sort, sense, wish, somewhat	797
Insight	Think, realize, question, accept, notice, perspective	259
Causation	Because, effect, hence, affect, based, create	135
Discrepancy	Should, would, could, hope, lack, odd	83
Tentative	Maybe, perhaps, guess, unknown, wonder	178
Certainty	Always, never, clearly, confident, commit, apparent	113
Differentiation	Hasn't, but, else, despite, although, except, exclude	81

LIWC is based on the understanding that people use words to outline how they view and interact with their world, and an analysis of words has great significance (Krieger et al., 2018). LIWC has been used in various contexts, including predicting final course grades (Robinson, Navea, & Ickes, 2013), influence within groups and group dynamics (Van Swol & Kane, 2018), studies on deception (Newman, Pennebaker, Berry, & Richards, 2003), and interactions between people (Kacewicz, Pennebaker, Davis, Jeon, & Graesser, 2014). The current version of LIWC has about 100 different categories which focus on topics like grammar, vocabulary usage, perceptions, emotions, social processes, and personal concerns (Bulkeley & Graves, 2018). Fast et al. (2017) assert that LIWC is now the standard-bearer for psychometrically validated categories.

4. MOOC pacing conditions

There are two primary methods of MOOC delivery: instructor-paced or self-paced. An instructor-paced MOOC uses a linear, sequential course structure dictated by the instructor. The self-paced, or on-demand, course has a more adaptive learning path, where the learner can decide in what order to view content, as the orientation to learning is student-centered. Table 2 summarizes how we expect the activity, role of a cohort and opportunities for knowledge scaffolding to compare between the pacing conditions.

Table 2. Comparison of expectations between pacing condition.

	Instructor-Paced	Self-Paced
Activity	More	Less
Cohort	More collaboration/community building	Less
Knowledge Scaffolding	More opportunities	Less opportunities

4.1. Activity

Sharif and Magrill (2015) predict that we would see more activity in the instructor-paced discussion forums and less in the self-paced forums. Overall, discussion forum participation is low (Onah, Sinclair, & Boyatt, 2014; Sharif & Magrill, 2015) and it will wane over time (Yang, 2014).

4.2. Cohort

Instructor-paced MOOCs, described by Sharif and Magrill (2015) as a cohort model, will have a defined start and end period contrasting with the self-paced MOOC that has a more open start and end period. Student engagement is linked to cognitive presence, and research has highlighted how critical the initial weeks are for engagement in MOOCs (Jordan, 2015; Perna et al., 2014). Since students in the instructor-paced course are following a similar schedule, there are greater changes for the type of interactions and engagement in the discussion forum that can foster cognitive presence.

In a self-paced course, students are not part of a cohort, but while the students may not be as engaged with each other, the self-paced MOOC can be more advantageous for self-directed learners if the content is of a high quality (Campbell et al., 2014). Campbell et al. suggest that a self-directed learner has less reliance on an instructor and therefore would likely prefer the self-paced structure for a MOOC.

4.3. Knowledge Scaffolding

Sharif and Magrill (2015) argue that a self-paced MOOC is too socially isolating and that the resulting disconnect damages the potential for "knowledge scaffolding." Scaffolding refers to the probing and follow-up questions that are posted in response to initial postings which further the discussion (Whipp, 2003). This is a legitimate concern, as even in the instructor-paced MOOCs, many students struggle to develop the social connections necessary to persist in the course (Yang et al., 2014). This can be exacerbated in a self-paced structure. The way students use a self-paced MOOC discussion forum impacts the way peers influence each other. This will differ from an instructor-paced course in that there will likely be more time between posts and replies. Students in a self-paced course may read a post made several weeks prior and offer their own insights. But the likelihood of the original post's author returning to the conversation and replying to their additions would be minimal. This lack of replies to one's contributions might be de-motivating. At the same time, by not being focused on specific deadlines, students in self-paced courses may be able to benefit from reading more peer posts before making their own.

5. Research aim and questions

Despite not being in a cohort model, we expected to find that the self-paced students would have higher average cognitive processing scores than their instructor-paced peers. We expected to see this because while the self-paced student may not post as frequently as their instructor-paced peer, their posts could still have high levels of cognitive processing benefiting from being able to read the other posts in the thread.

Also, a self-paced student who focused on longer threads would gain more opportunities to deeply engage with the content, which Hecking, Hoppe, and Harrer (2015) hypothesize would result in higher learning gains. We also wanted to provide empirical evidence to resolve the disagreement between two findings. Campbell, Gibbs, Najafi, and Severinski (2014) found that students in self-paced MOOCs were still actively engaged in discussion forums despite not having a defined cohort. These findings contradict Sharif and Magrill (2015), who suggest that the opposite would occur—students would feel isolated in the self-paced MOOC and, therefore, not engage in knowledge construction. To further explore the relationship between the pacing condition and cognitive processing, this study answers the following questions:

1. Is there a difference in cognitive processing between students who participate in different pacing conditions of a MOOC?
2. What characteristics (e.g. learner demographics and engagement with course content, linguistic qualities of post, post classification) predict cognitive processing in discussion forum posts?

6. Method

6.1 Participants

The participants for this study made discussion forum posts in either a self-paced or instructor-paced section of the MOOC *Visualizing Japan (1850s-1930s):Westernization, Protest Modernity* offered by HarvardX on the edX platform. Participants completed an optional pre-course survey that asked demographic questions including their level of education and gender. Through a data use agreement with the Harvard VPAL Research Team, this identifiable data was provided to the researchers for the study. For the two versions of this course, 11,216 students registered, while this study focuses only on the 671 students who contributed to the discussion forums. Of these contributors, approximately 42% (approximately 45% in the self-paced and approximately 39% in the instructor-paced) were “viewers” and accessed at least one course chapter, approximately 15% were “explorers” (approximately 15% in both self-paced and instructor-paced courses) and accessed at least half of the course chapters, and approximately 43% were “completers” (approximately 40% in the self-paced and approximately 46% in the instructor-paced) and completed all of the course requirements (Table 3). The study participants were highly educated as 41% had either a bachelor’s or associates degree, 30% had a masters, and 5% had a doctorate. More than half (53%) of the participants were female.

Table 3. Descriptive statistics showing students by course engagement status.

	Self-Paced (n = 353)	Instructor-Paced (n = 318)
# of students who were viewers	159	123
# of students who were explorers	53	48
# of students who were completers	141	147

6.2 Context

The *Visualizing Japan (1850s-1930s): Westernization, Protest Modernity* MOOC is a collaboration between MITx and Harvardx and is organized into four modules. Both the self-paced and instructor-paced had identical learning objectives and content, including discussion forum prompts and assessments. The only difference between the courses was how long students had to engage with the content and each other within the course. The instructor-paced version was designed to take six weeks and require between three and five hours of effort by each learner weekly and is in the art and culture subject area. The instructor-paced version used in this study ran from 09/15/15 to 11/3/15 and the self-paced version ran from 09/1/16 to 09/31/17. At the end of the instructor-paced course, the content was used to create the self-paced version. This study was done as a post-hoc analysis – meaning that we received the identifiable data but did not have direct contact with the course designers or any of the students.

6.3 Procedure

The initial dataset included 3,851 posts from both instructors and students. Using the procedure outlined in Figure 1, we narrowed our dataset to 2,423 posts from 671 unique students. For our LIWC analysis, researchers uploaded a .csv file, with each discussion forum post occupying its own row, and LIWC appended the different values. The LIWC-generated variables included the number of words, number of six-letter words, number of dictionary words, and the summary variables of analytical thinking, tone, authenticity, and clout, as well as the cognitive processing score. We were interested in exploring which linguistic properties influenced cognitive processing, e.g. the length of a post or specific types of words being used, and used these variables to explore these potential relationships.

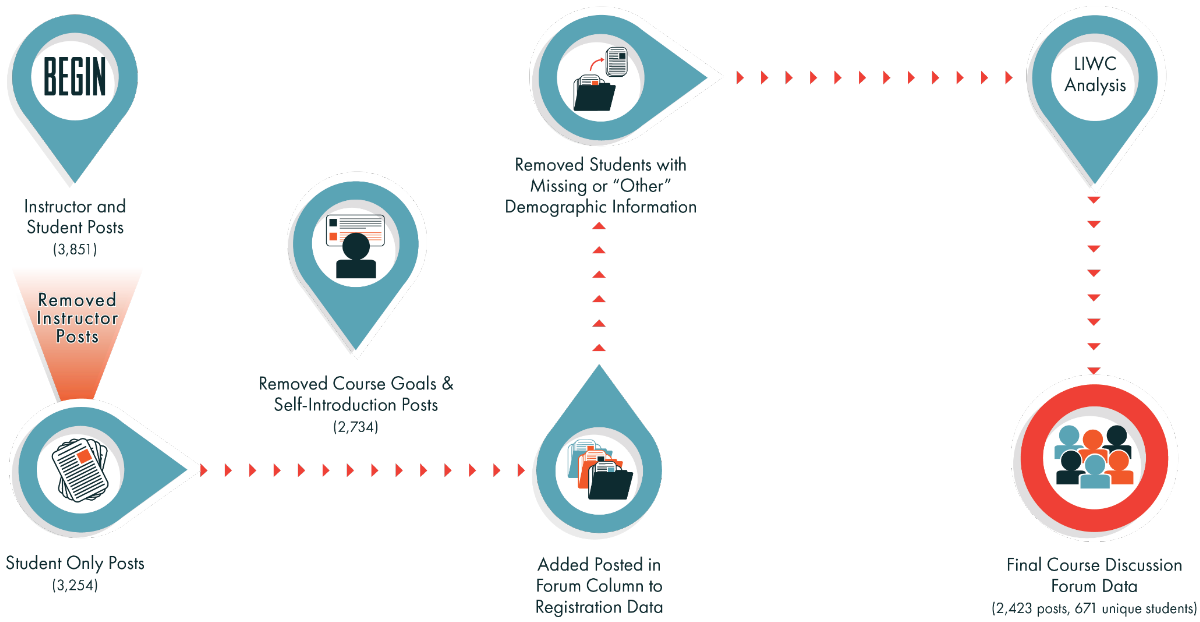

Figure 1. Data Clean-Up Procedure.

6.4 Variables

The dependent variable for this study is cognitive processing, as calculated by LIWC. LIWC also generated values for tone, authenticity, number of six-letter words, word count, and words per sentence. A bivariate correlation showed that none of these were significantly correlated to cognitive processing scores. Means and standard deviations of posting features used in the model appear in Table 4. The values for cognitive processing (cogproc), three summary variables of analytical thinking (analytic), clout (clout), and authenticity (authentic) are reported as percentage of words while the dictionary value is the number of words used.

Table 4. Means and Standard Deviations of Posting Features

		Cogproc	*Analytic*	*Clout*	*Authentic*	*Dictionary*
Pacing	Self-Paced	11.98	76.28	52.99	48.13	82.63
	(*n* = 353)	(5.80)	(19.54)	(18.36)	(25.65)	(9.36)
	Instructor-	12.05	76.68	54.47	44.25	82.30
	Paced	(5.35)	(21.12)	(19.38)	(25.56)	(9.07)
	(*n* = 318)					
Gender	Male	12.29	75.98	54.48	46.48	82.40
	(*n* = 314)	(5.54)	(21.02)	(18.40)	(25.75)	(9.19)
	Female	11.77	76.90	52.99	46.12	82.55
	(*n* = 357)	(5.63)	(19.65)	(19.24)	(25.62)	(9.25)
Education	High School	12.03	73.99	55.45	46.39	82.33
Background	(*n* = 160)	(5.87)	(22.76)	(20.14)	(25.39)	(10.70)
	College	11.98	77.50	53.68	47.59	82.62
	(*n* = 275)	(4.84)	(18.52)	(17.39)	(24.45)	(9.66)
	Master	12.08	76.95	51.86	45.03	82.20
	(*n* = 203)	(6.10)	(20.26)	(19.60)	(26.67)	(7.56)
	Doctorate	11.85	77.02	56.49	42.74	83.71
	(*n* = 33)	(6.89)	(21.87)	(19.12)	(30.57)	(6.93)
Type of	Initial	12.34	76.19	53.33	41.15	83.11
Posts	(*n* = 353)	(6.77)	(25.39)	(23.64)	(32.48)	(8.57)

	Response	12.65	75.04	53.04	41.26	83.17
	(n = 353)	(7.62)	(26.50)	(23.38)	(30.89)	(11.29)
	Comment	13.59	70.55	51.22	42.61	84.88
	(n = 353)	(7.95)	(28.14)	(25.22)	(33.32)	(8.82)
Status	Viewed	11.41	79.56	54.58	52.49	81.94
	(n = 282)	(6.51)	(20.80)	(20.63)	(29.81)	(9.07)
	Explored	11.91	79.15	56.05	41.09	80.91
	(n = 101)	(5.75)	(19.55)	(19.19)	(23.75)	(12.75)
	Completed	12.64	72.52	51.98	42.05	83.55
	(n = 288)	(4.38)	(19.41)	(16.69)	(20.13)	(7.67)

Note. Numbers in parentheses are standard deviations.

Although the means in Table 5 vary by groups classified by gender and educational background, we cannot conclude that one group's cognitive processing scores are higher than the other group because our null hypothesis is that there is not a difference (the difference is zero) and we failed to reject the null hypothesis.

6.5 Data Analytical Procedure

As each student has multiple posts and the posts are not independent with each other, use of tradtional statistical procedure violates the assumption of independent observation. Means and standard deviations of post type and word count per post appear in Table 5.

Table 5. Means and Standard Deviations of post type and word count per post

	Self-Paced (n = 353)		Instructor-Paced (n = 318)	
	M	*SD*	*M*	*SD*
No. of Initial Posts per student	3.09	3.12	2.16	1.81
No. of Response Posts per student	2.01	2.82	1.95	2.20
No. of Comment Posts per student	0.77	1.24	0.99	1.56
Average Word Counts per Post	53.36	39.76	60.38	48.52

A random intercept two-level hiearchical linear model (HLM2) was employed to account for the nesting feature of data: multiple postings within each student. The null model (unconditional model without any predictors) was used as the baseline model to calculate the effect size of the full (conditional) model (Raudenbush & Bryk, 2002). The use of HLM is a novel approach to analysis of MOOC discussion forum data, made possible because of the identifiable data provided by Harvard. This allowed us to have two levels of data with all posts and responses nested within each individual student, which is a more robust approach than a multiple linear regression approach because the violation of independent observation within each individual student was accounted for and that the between-individual and within-individual variances were included in the same model. The two-level full HLM is as follows:

Level 1: Response Level

$$COGPROC_{ij} = \beta_{0j} + \beta_{1j}*(RESPONSE_{ij}) + \beta_{2j}*(COMMENT_{ij}) + \beta_{3j}*(ANALYTIC_{ij}) + \beta_{4j}*(CLOUT_{ij}) + \beta_{5j}*(AUTHENTI_{ij}) + \beta_{6j}*(DIC_{ij}) + r_{ij}$$

Level 2: Individual Student Level

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(EXPLORED_j) + \gamma_{02}*(CERTIFIE_j) + \gamma_{03}*(COMPLETE_j) + \gamma_{04}*(COLLEGE_j) + \gamma_{05}*(MASTERS_j) + \gamma_{06}*(DOCTORAT_j) + \gamma_{07}*(GENDER_j) + \gamma_{08}*(COURSE_j) + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{2j} = \gamma_{20}$$

$$\beta_{3j} = \gamma_{30} + \gamma_{31}*(COURSE_j)$$

$$\beta_{4j} = \gamma_{40} + \gamma_{41}*(COURSE_j)$$

$$\beta_{5j} = \gamma_{50} + \gamma_{51}*(COURSE_j)$$

$$\beta_{6j} = \gamma_{60} + \gamma_{61}*(COURSE_j)$$

Level 1 of the model is the response level. The dependent variable in Level 1 regression is the cognitive processing score of the postings, and independent variables in Level 1 regression include the dummy coded variables of response and comment (initial posting was used as the baseline comparison). Other independent variables in the Level 1 regression include features of the posting (i.e., analytical, clout, authenticity, and use of dictionary words). All variables in Level 1 are characteristics of the postings. The intercept in this Level 1 regression represents the estimated mean cognitive processing score of initial posts.

The intercept and slopes in the Level 1 regression became dependent variables in the Level 2 regressions. The first regression in Level 2, the individual student level, predicts the intercept in Level 1 with student characteristics such as educational background (college, master's, doctorate degrees) using high school graduates as the reference group. Also included in this regression are student gender (male as the reference group), course (self-paced as the reference group), and the outcome of the student's course engagement (explored, certificate, completed). All of these Level 2 independent variables were dummy coded so that the slopes represent the differences between the reference group and the focal group. Course (self-paced versus instructor-paced) was also included in the regressions in Level 2 to examine the moderating effect of course on the relationships between cognitive processing and features of the posts (i.e., analytical, clout, authenticity, and use of dictionary words).

7. Findings

In this study, we explored student participation in self-paced versus instructor-paced discussion forums and whether this pacing condition influenced the demonstration of cognitive processing. We expected to find that students in the self-paced courses would have higher average cognitive processing scores than students in the instructor-paced courses. This may seem counterintuitive because one would assume that more interactions between students would occur in the instructor-paced courses vs the self-paced course where students could be making posts to a ‘dead’ forum and not receiving replies. We further expected that there would still be activity within the self-paced discussion forums (Campbell et al., 2014) and the additional time and consideration of the posts would lead to more opportunities for students to engage in knowledge construction with each other which in turn would be reflected in higher levels of cognitive presence (Kent et al., 2016). This expectation is supported by Kanuka and Garrison (2004) who posit that cognitive presence requires sustained communication and student-student and student-content interaction.

Results from the HLM are presented in Table 6. The HLM explained 15% of the variance in the cognitive processing of the postings. The results of a HLM showed that the strongest predictor of cognitive processing were the number of words that appeared in the dictionary. With regard to RQ1, we did not find that the course (self-paced versus instructor-paced) significantly influenced the cognitive processing of the posts. With regard to RQ1, we found that none of the student background information (gender, education) or engagement with the course content (explored or certified) significantly influenced cognitive processing in the postings. No statistically significant relationship was noted between cognitive processing of the posting and the authenticity summary variable score of the postings. We did find some relationships when exploring the linguistic qualities of the posts and cognitive processing. Although this effect size is small (Cohen, 1988), the model showed statistically significant positive relationships between cognitive processing of the posting and use of dictionary words (DIC). This would suggest that students were using formal language (use of dictionary words). Additionally, a negative relationship between cognitive processing and the analytical summary variable and clout summary variable was found. This suggests that there are aspects of analytical thinking and clout that are not supportive of cognitive processing.

For instance, if the posts were demonstrating higher levels of tentativeness or discrepancy (two aspects of cognitive processing), this could result in lower scores of analytical thinking. Additionally, expressing thoughts that showed confidence and certainty (clout) may not be as supportive or fostering the type of ongoing discussion and discourse necessary for cognitive processing.

Table 6. Fixed Effects Estimates (Top) and Variance-Covariance Estimates (Bottom) for the HLM Predictors of Cognitive Processing

Parameter	Coefficient	*t*	*df*	*p*
Level 1: Response Level				
Response	0.17	0.42	1742	.68
Comment	0.63	1.33	1742	.18
Analytic	-0.07	-5.85	1742	< .001
Course (moderate)	0.03	1.55	1742	.12
Clout	-0.06	-5.83	1742	< .001
Course (moderate)	0.02	0.91	1742	.36
Authentic	0.01	0.79	1742	.43
Course (moderate)	0.02	1.37	1742	.17
DIC	0.16	4.04	1742	.< .001
Course (moderate)	0.03	0.46	1742	.65
Level 2: Individual Student Level				
Explored	0.60	1.08	662	.28
Certified	0.78	1.46	662	.15
Completed	0.24	0.40	662	.69
College Degree	-0.27	-0.64	662	.53
Master’s Degree	-0.62	-1.30	662	.19

Doctorate Degree	-0.74	-0.81	662	.42
Gender (Female)	0.18	0.57	662	.57
Course (lecture)	-0.17	-0.40	662	.69

6.1. Dictionary Words

The use of dictionary words was the only variable that we found to be positively associated with cognitive processing. Cognitive processing reflects a level of language complexity (Van Swol, Prahl, Kolb, Lewis, & Carlson, 2016) and further demonstrates the level to which the writer has organized their thoughts (Cohn, Mehl, & Pennebaker, 2004). This is supported by Bulkeley and Graves (2018) who assert that the higher the number of dictionary words used, the more use of formal language in the post. Therefore, a recommendation for instructors interested in fostering cognitive processing, should encourage students to more formal language in their writing.

6.2. Clout

Clout is one of the four summary variables in LIWC meant to measure the level of confidence and certainty in the post (Pennebaker et al., 2015) and the algorithm was developed based on results of studies focusing on personal interactions (Kacewicz et al., 2014). O'Dea, Larsen, Batterham, Calear, and Christensen (2017) emphasize that the clout summary variable reflects the social status, confidence or leadership that individuals demonstrate through their writing expression. Jordan & Pennebaker (2017) further define clout as having a higher number of 'we'-words and social words and a lower number of 'I'-words and negations. We found that this was negatively associated with cognitive processing. This could be due to the fact that showing more confidence may not foster as much discussion and openness for discussion.

6.2. Analytical Thinking

Analytical thinking is one of four global high-level text properties in LIWC (Simms et al., 2017) and is one of the summary variables output by LIWC. This summary variable measures the degree to which a post contains words that demonstrate formal, logical, and hierarchical thinking patterns—something that is a component of cognitive presence (Tausczik & Pennebaker, 2010). The fact that this variable was negatively associated with the cognitive processing score indicates that there may be underlying components of cognitive processing that were in contradiction with the analytical score. Unfortunately, like the clout summary variable, the analytical thinking score is a non-transparent dimension which prevents an in-depth examination of the words used to calculate this score.

8. Discussion

The results of this study provide more insight on engagement in self-paced MOOC forums as well as the utility of learning analytics for understanding student activity within MOOCs. It also identifies areas for future exploration.

8.1 Engagement in the Self-Paced Forums

This study showed that students were making substantive posts within the self-paced versions, supporting the findings of Campbell et al. (2014). If the content and discussion prompts are well designed, it can create opportunities for students to make substantive posts in the forum. It also suggests that students' decision to post, and their demonstration of cognitive processing, is more predicted by factors other than the pacing condition. These factors can be their interest in topic, motivation to complete the assignments, and peer interactions. Course designers should carefully consider how the discussion forum is structured through pre-determined threads or categories and compelling prompts that can foster critical thinking even in a self-paced course. While this study used a post-hoc analysis so we were not able to interview the instructors, we can see that the prompts created in the two versions did foster high levels of discussion. Future studies that can incorporate interviews with course designers to learn their intention for the forums would add another level to the analysis of the discussion forums. The findings of this study also suggest that self-paced forums have the potential to provide the student-to-student and student-to-content interactions that are important in an online learning environment.

8.2 Utility of learning analytics

The nature of MOOCs represents both a challenge and a benefit for educational research purposes. Large enrollments result in extensive datasets and multiple data points, such as student log files of interactions, discussion forum posts, and activity completion scores. This vast quantity of data presents a unique opportunity to explore connections between student engagement, learning goals, and outcomes (Henrie, Halverson, & Graham, 2015). Through text analysis and clickstream data, we extend existing research to examine the impact of course structure on learner behaviors within MOOCs. Our research contributes not only to the literature on the development of cognitive presence in MOOC courses, but also explores two distinct course structures—instructor-paced and self-paced—and their influence on this development. Furthermore, analysis of multiple versions of the same course will address the literature gaps on differences in learner behaviors and levels of engagement within the same MOOC course (Gallagher & Savage, 2016). As Touati (2016) notes, MOOC instructors face the challenge of creating learning environments that will allow for personalized interactions in situations where there are far more students than available instructors. Thus, personalized interactions need to occur at the student level, and understanding the influence of

MOOC structure on these types of interactions is critical for studying and improving MOOCs.

8.3 Future Exploration

Through our analysis, we were able to take a large corpus and begin to gain insight into understanding the quality of posts in discussion forums. While our findings showed that educational background, gender and pacing condition may not have significant influence on cognitive processing, we did identify three specific elements that do – the number of dictionary of words and the clout and analytical thinking summary variables. The summary variables are non-transparent, and we need to better understand how these two variables are negatively impacting cognitive processing. The cognitive processing category is made up of subscores, such as insight and causation, and by using these subscores we may be able to gain even more insight into how these interact with the summary variables. Through this subsequent analysis we will be able to make more refined recommendations for instructors.

9. Conclusion

Ultimately we found that the pacing condition, gender or educational background were not significant predictors of cognitive processing. Based on our findings, it does not appear that cognitive processing is influenced by whether a course is self-paced or instructor-paced. This paper presented an examination of the influence that the pacing condition has on the demonstration of cognitive processing in MOOCs. Through an analysis of the forums, we furthered the understanding of student participation within online and open learning environments, such as MOOCs (Chiu & Hew, 2017). This study compared the two iterations in terms of forum engagement and learning as measured by cognitive processing (outcome measure from the Linguistic Inquiry Word

Count (LIWC) tool). We focused on the student posts within each course discussion forum to determine whether there is a statistically significant difference in the demonstration of cognitive processing between the instructor-paced and self-paced course participants. The results of a hierarchical linear model showed that the strongest predictors of cognitive processing were the number of words that appeared in the dictionary. The negative associations of clout and analytical thinking on cognitive processing will frame future research as we seek to gain a deeper understanding of student learning and engaging in MOOC discussion forums.

References

Adams, C., Yin, Y., Vargas Madriz, L. F., & Mullen, C. S. (2014). A phenomenology of learning large: the tutorial sphere of xMOOC video lectures. *Distance Education*, *35*(2), 202–216. https://doi.org/10.1080/01587919.2014.917701

Akyol, Z., Arbaugh, J. B., Cleveland-Innes, M., Garrison, D. R., Ice, P., Richardson, J. C., & Swan, K. (2009). A response to the review of the community of inquiry framework. *International Journal of E-Learning & Distance Education*, *23*(2), 123–136.

Akyol, Z., & Garrison, D. R. (2011). Understanding cognitive presence in an online and blended community of inquiry: Assessing outcomes and processes for deep approaches to learning. *British Journal of Educational Technology*, *42*(2), 233–250. https://doi.org/10.1111/j.1467-8535.2009.01029.x

Al-Imarah, A. A., & Shields, R. (2018). MOOCs, disruptive innovation and the future of higher education: A conceptual analysis. *Innovations in Education and Teaching International*, *3297*, 1–12. https://doi.org/10.1080/14703297.2018.1443828

Beckmann, J., & Weber, P. (2016). Cognitive presence in virtual collaborative learning. *Interactive Technology and Smart Education*, *13*(1), 52–70. https://doi.org/10.1108/ITSE-12-2015-0034

Bonk, C. J., Lee, M. M., Reeves, T. C., & Reynolds, T. H. (2017). The emergence and design of massive open online courses (MOOCs). In R. A. Reiser & J. V. Demsey (Eds.), *Trends and issues in instructional design and technology* (4th editio, pp. 250–258). Boston, MA: Pearson.

Breivik, J. (2016). Critical thinking in online educational discussions measured as progress through inquiry phases: A discussion of the cognitive presence construct in the community of inquiry framework. *International Journal of E-Learning & Distance Education*, *32*(1), 1–16.

Bulkeley, K., & Graves, M. (2018). Using the LIWC program to study dreams. *Dreaming*, *28*(1), 43–58. https://doi.org/10.1037/drm0000071

Campbell, J., Gibbs, A., Najafi, H., & Severinski, C. (2014). A comparison of learner intent and behaviour in live and archived MOOCs. *International Review of Research in Open and Distance Learning*, *15*(5), 235–262. https://doi.org/10.19173/irrodl.v15i5.1854

Chiu, T. K. F., & Hew, T. K. F. (2017). Factors influencing peer learning and performance in MOOC asynchronous online discussion forum. *Australasian Journal of Educational Technology*, *34*(4), 16–28. https://doi.org/10.14742/ajet.3240

Cohen, A., & Holstein, S. (2018). Analysing successful massive open online courses using the community of inquiry model as perceived by students. *Journal of Computer Assisted Learning*, (January), 1–13. https://doi.org/10.1111/jcal.12259

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science*, *15*(10), 687–693. https://doi.org/10.1111/j.0956-7976.2004.00741.x

Donohue, W. A., Liang, Y., & Druckman, D. (2014). Validating LIWC dictionaries: The Oslo I Accords. *Journal of Language and Social Psychology*, *33*(3), 282–301. https://doi.org/10.1177/0261927X13512485

Dowell, N., & Graesser, A. C. (2014). Modeling learners' cognitive, affective, and social processes through language and discourse. *Journal of Learning Analytics*, *1*(3), 183–186. Retrieved from http://epress.lib.uts.edu.au/journals/index.php/JLA/article/view/4203

Fast, E., Chen, B., & Bernstein, M. S. (2017). Lexicons on demand: Neural word embeddings for large-scale text analysis. In *Proceedings of the Twenty-Sixth International Joint Conference on Artifical Intelligence (IJCAI-17)* (pp. 4836–4840).

Gallagher, S. E., & Savage, T. (2016). Comparing learner community behavior in multiple presentations of a massive open online course. *Journal of Computing in Higher Education*, *28*(3), 358–369. https://doi.org/10.1007/s12528-016-9124-y

Garrison, D. R. (2007). Online community of inquiry review: social, cognitive, and teaching presence issues. *Journal of Asynchronous Learning Networks*, *11*(1), 61–72. https://doi.org/10.1128/JB.05513-11

Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education*, *15*(1), 7–23. https://doi.org/10.1080/08923640109527071

Hara, N., Bonk, C. J., & Anjeli, C. (2000). Content analysis of online discussions in an applied educational pyschology course. *Instructional Science*, *28*, 115–152.

Hayes, S. (2015). *MOOCs and quality: A review of the recent literature*.

Hecking, T., Hoppe, H. U., & Harrer, A. (2015). Uncovering the structure of knowledge exchange in a MOOC discussion forum. *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015 - ASONAM '15*, 1614–1615. https://doi.org/10.1145/2808797.2809359

Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement in technology-mediated learning: A review. *Computers & Education*, *90*, 36–53.

Hew, K. F., & Cheung, W. S. (2014). Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges. *Educational Research*

Review, *12*, 45–58. https://doi.org/10.1016/j.edurev.2014.05.001

Humphreys, A., & Wang, R. J.-H. (2018). Automated text analysis for consumer research. *Journal of Consumer Research*, *44*(6), 1274–1306. https://doi.org/10.1093/jcr/ucx104

Jordan, K. (2015). Massive open online course completion rates revisited: Assessment, length and attrition. *International Review of Research in Open and Distributed Learning*, *16*(3), 341–358. https://doi.org/10.13140/RG.2.1.2119.6963

Jordan, K. N., & Pennebaker, J. W. (2017). Trump's first state of the union address. Retrieved August 18, 2018, from https://wordwatchers.wordpress.com/2017/03/01/trumps-first-state-of-the-union-address

Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology*, *33*(2), 125–143. https://doi.org/10.1177/0261927X13502654

Kahn, J. H., Tobin, R. M., Massey, A. E., & Anderson, J. A. (2007). Measuring emotional expression with the linguistic inquiry and word count. *The American Journal of Psychology*, *120*(2), 263–286. https://doi.org/10.2307/20445398

Kanuka, H., & Garrison, D. R. (2004). Cognitive presence in online learning. *Journal of Computing in Higher Education*, *15*(2), 21–39. https://doi.org/10.1007/BF02940928

Kent, C., Laslo, E., & Rafaeli, S. (2016). Interactivity in online discussions and learning outcomes. *Computers & Education*, *97*, 116–128. https://doi.org/10.1016/j.compedu.2016.03.002

Khazaei, T., Lu, X., & Mercer, R. (2017). Writing to persuade: Analysis and detection of persuasive discourse. In *Proceedings of the iConference 2017* (pp. 203–215).

https://doi.org/10.9776/17022

Kilis, S., & Yıldırım, Z. (2018). Investigation of community of inquiry framework in regard to self-regulation, metacognition and motivation. *Computers & Education*, *126*(June), 53–64. https://doi.org/10.1016/j.compedu.2018.06.032

Kovanović, V., Joksimović, S., Waters, Z., Gašević, D., Kitto, K., Hatala, M., & Siemens, G. (2016). Towards automated content analysis of discussion transcripts: A cognitive presence case. *Learning Analytics and Knowledge (LAK'16)*, 15–24. https://doi.org/10.1145/2883851.2883950

Krieger, K., Watkins, P. Lou, Gerber, M. R., Pham, H., & Bauman, L. (2018). Weigh your words: An exploration of natural word use in a fat studies course. *Fat Studies*, *7*(1), 56–68. https://doi.org/10.1080/21604851.2017.1361280

Lowenthal, P., Snelson, C., & Perkins, R. (2018). Teaching massive, open, online, courses (MOOCs): Tales from the front line. *The International Review of Research in Open and Distributed Learning*, *19*(3), 1–19. Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/3505/4660

Major, C. H., & Blackmon, S. J. (2016). Massive open online courses: Variations on a new instructional form. *New Directions for Institutional Research*, *167*, 11–25. https://doi.org/10.1002/ir.20151

McLoughlin, D., & Mynard, J. (2009). An analysis of higher order thinking in online discussions. *Innovations in Education and Teaching International*, *46*(2), 147–160. https://doi.org/10.1080/14703290902843778

Meyer, K. A. (2003). Face-to-face versus threaded discussions: The role of time and higher-order thinking. *Journal of Asynchronous Learning Networks*, *7*(3), 55–65.

Moore, R. L. (2019). The role of data analytics in education: Possiblities and limitations. In B. Khan, J. Corbeil, & M. Corbeil (Eds.), *Responsible Analytics and*

Data Mining in Education: Global Perspectives on Quality, Support, and Decision Making (pp. 101–118). New York: Routledge.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, *29*(5), 665–675.

O'Dea, B., Larsen, M. E., Batterham, P. J., Calear, A. L., & Christensen, H. (2017). A linguistic analysis of suicide-related Twitter posts. *Crisis*, *38*(5), 319–329. https://doi.org/10.1027/0227-5910/a000443

Onah, D. F. ., Sinclair, J., & Boyatt, R. (2014). Exploring the use of MOOC discussion forums. In *Proceedings of the London International Conference on Education (LICE-2014)* (pp. 1–4). London.

Oztok, M., Zingaro, D., Brett, C., & Hewitt, J. (2013). Exploring asynchronous and synchronous tool use in online courses. *Computers & Education*, *60*(1), 87–94. https://doi.org/10.1016/j.compedu.2012.08.007

Pennebaker, J. W. (2011). *The secret life of pronouns: What our words say about us*. New York, NY: Bloomsbury Press.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015. *Austin, TX: University of Texas at Austin*. Austin, TX: University of Texas at Austin. Retrieved from https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf

Perna, L. W., Ruby, A., Boruch, R. F., Wang, N., Scull, J., Ahmad, S., & Evans, C. (2014). Moving through MOOCs: Understanding the Progression of Users in Massive Open Online Courses. *Educational Researcher*, *43*(9), 421–432. https://doi.org/10.3102/0013189X14562423

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Robinson, R. L., Navea, R., & Ickes, W. (2013). Predicting final course performance from students' written self-introductions: A LIWC analysis. *Journal of Language and Social Psychology*, *32*(4), 469–479. https://doi.org/10.1177/0261927X13476869

Sharif, A., & Magrill, B. (2015). Discussion forums in MOOCs. *International Journal of Learnings, Teaching and Educational Research*, *12*(1), 119–132.

Shea, P., & Bidjerano, T. (2010). Learning presence: Towards a theory of self-efficacy, self-regulation, and the development of a communities of inquiry in online and blended learning environments. *Computers & Education*, *55*(4), 1721–1731.

Siemens, G. (2005). A learning theory for the digital age. *Elearnspace Everything Learning*, *2*(1), 1–8. Retrieved from http://www.ingedewaard.net/papers/connectivism/2005_siemens_ALearningTheoryForTheDigitalAge.pdf

Simms, T., Ramstedt, C., Rich, M., Richards, M., Martinez, T., & Giraud-Carrier, C. (2017). Detecting cognitive distortions through machine learning text analytics. In *2017 IEEE International Conference on Healthcare Informatics (ICHI)* (pp. 508–512). IEEE. https://doi.org/10.1109/ICHI.2017.39

Slotter, E. B., & Ward, D. E. (2015). Finding the silver lining: The relative roles of redemptive narratives and cognitive reappraisal in individuals' emotional distress after the end of a romantic relationship. *Journal of Social and Personal Relationships*, *32*(6), 737–756. https://doi.org/10.1177/0265407514546978

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social*

Psychology, *29*(1), 24–54. https://doi.org/10.1177/0261927X09351676

Touati, A. (2016). Self-directed learning in MOOCs: A disconnect between theory and practice. *Middle Eastern & African Journal of Educational Research*, (19), 15–30.

Van Swol, L. M., & Kane, A. A. (2018). Language and group processes: An integrative, interdisciplinary review. *Small Group Research*, 1–36. https://doi.org/10.1177/1046496418785019

Van Swol, L. M., Prahl, A., Kolb, M. R., Lewis, E. A., & Carlson, C. (2016). The language of extremity: The language of extreme members and how the presence of extremity affects group discussion. *Journal of Language and Social Psychology*, *35*(6), 603–627. https://doi.org/10.1177/0261927X16629788

Veletsianos, G., & Shepherdson, P. (2016). A systematic analysis and synthesis of the empirical MOOC literature published in 2013-2015. *International Review of Research in Open and Distributed Learning*, *17*(2), 198–221. https://doi.org/10.19173/irrodl.v17i2.2448

Wang, Q., Woo, H. L., & Zhao, J. (2009). Investigating critical thinking and knowledge construction in an interactive learning environment. *Interactive Learning Environments*, *17*(1), 95–104.

Whipp, J. (2003). Scaffolding critical reflection in online discussions: Helping prospective teachers think deeply about field experiences in urban schools. *Journal of Teacher Education*, *54*, 321–333.

Wong, J.-S., Pursel, B. K., Divinsky, A., & Jansen, B. J. (2015). An analysis of MOOC discussion forum interactions from the most active users. In N. Agarwal, K. Xu, & N. Osgood (Eds.), *Social Computing, Behavioral-Cultural Modeling, and Prediction* (pp. 452–457). Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-16268-3_58

Yang, D., Wen, M., Kumar, A., Xing, E. P., & Rosé, C. P. (2014). Towards an integration of text and graph clustering methods as a lens for studying social interaction in MOOCs. *International Review of Research in Open and Distance Learning*, *15*(5), 215–234.

Yang, Q. (2014). Students motivation in asynchronous online discussions with MOOC Mode. *American Journal of Educational Research*, *2*(5), 325–330.