Identifying deep structural features of MOOC lectures and their impact on learning outcomes: A nonlinear dynamics approach

Yohei Kato: S2843560
Department of Educational Sciences, Evidence-Based Education
University of Groningen
First Supervisor: Dr. Michael Tscholl
Second Supervisor: Dr. Filitsa Dingyloudi
Date: July 15th 2018
Place: Groningen

"I, Yohei Kato, hereby confirm that I completed this master's thesis entitled "Identifying deep structural features of MOOC lectures and their impact on learning outcomes: A nonlinear dynamics approach" independently, and that I have not heretofore presented the contents of this thesis in any other shape or form as (partial or complete) fulfillment for any degree or course to another department or university, and I have listed all references used, and have given credit to all additional sources of assistance."

Abstract

The purpose of this study was to examine the impact of deep structural features of MOOC lectures (available as video and text) on outcome measures, specifically, completion rates, quiz scores, and test scores. We applied standardized dispersion analysis to detect variability scales of common formal features (sentence length, the number of stopwords, and the number of domain-specific words) of the lectures. A surface analysis of the features showed no significant correlation between them, but the variability scales correlated with outcome measures: (1) the variability scale of domain-specific words has a strong positive correlation with the completion rates, (2) the variability scale of stopwords has a strong negative correlation with the quiz scores, and (3) the variability scale of stopwords has a strong positive correlation with the test scores. A simple regression analysis showed that: (1) the variability scale of domain-specific words predicts the completion rates, (2) the variability scale of stopwords predicts the quiz and test scores. Through a multiple regression analysis to identify the relative weight of all three metrics, we identified three models that predict the quiz scores. Our findings contribute to the question on how to characterize online lectures as well as how to design them for effective learning.

Keywords: MOOC, nonlinear dynamics, variability scale, standardized dispersion analysis, learning outcomes

Word count: (7977/8000)

Introduction

Online lectures are increasingly used by people with constraints of time and space to improve their competencies and knowledge. The most significant development in this respect is the emergence of massive open online courses (MOOCs). Previous systematic review (Veletsianos & Shepherdson, 2016) on MOOCs found out some important characteristics of past MOOC studies. One is that the majority of studies just use descriptive statistics and correlational analysis because they are convenient and efficient to use. In other words, most previous studies on MOOCs just apply a simple analytical method. Another finding is that more than 80% of studies focus just on learners, whereas 10.9% of studies examine context and impact of MOOCs. For instance, the former learner-focused studies examine motivation types, learning experiences, drop-out rates,

and demographics, while the latter studies address the usefulness of MOOCs as an educational medium and their economic impact. In sum, the first finding points out a methodological limitation of previous MOOC studies, whereas the second finding shows a limitation of research scope. Even though some studies focus on the context and effect of MOOCs, their research scope is broad in that they do not thoroughly examine the effect of MOOC contents on learning.

A different systematic study (Kennedy, 2014) shows that few studies on MOOCs have examined the effects of MOOC lectures on learning outcomes. This systematic review reveals that even if studies focus on MOOC lectures, they have a tendency to just examine learners' views or experiences about the MOOC course rather than investigate more deeply the impact of the course contents on learning outcomes.

Research in the field of MOOCs is partially hampered by the use of traditional methodologies (e.g., descriptive statistics, correlational analysis, etc.) which are usually inadequate to capture the dynamics of lectures (Veletsianos & Shepherdson, 2016). Our work is motivated by the assumption that lectures should be considered as a "composition," similar to a piece of music or a novel. While some research has applied non-traditional methodologies to identify the dynamic nature of composition in music and literature (Mainzer, 2009; González-Espinoza, Larralde, Martínez-Mekler, & Müller, 2017; Glattre, & Glattre, 2018), there is as yet very little work applying these methods on online lectures. Prior research on the impact of formal features of lectures on learning outcomes has not produced conclusive findings (i.e., Auble & Franks,1983; Haas & Losee, 1994; Mikk, 2008).

This study is especially concerned with one deep structural feature: variability. This work will identify the variability of three formal features of the transcripts of video lectures: (1) sentence length, (2) stopwords, and (3) domain-specific words. In contrast to many other studies, the unit of analysis of this study extends to the lecture as a whole deriving a single measure, characteristic for a lecture; specifically, this study applies a nonlinear dynamics approach to detect not only the presence and absence of formal features, but how their presence varies over the course of a lecture by examining the time-series data of the transcripts. The effect of various forms of variability (defined below) will be examined in relation to outcome measures; completion rates (the ratio of

total learners to learners indicating that they have completed viewing a lecture video), quiz scores, and test scores.

Overview of Variability in Learning and Previous Studies

The essential meaning of variability is that any changeable phenomena involve intrinsic fluctuations (Kunnen & van Geert, 2012; van Dijk & van Geert, 2007). For example, lectures sometimes include a difficult concept and sometimes an easy concept. In a word, variability represents actual changes in learning (Steenbeek & van Geert, 2007b).

Our work regards the variability of lecture videos as learning inputs for learners. Recent studies (i.e., Drożdż, Oświęcimka, Kulig, Kwapień, Bazarnik, Grabska-Gradzińska, & Stanuszek, 2016; Meindertsma, van Dijk, Steenbeek, & van Geert, 2013; Meindertsma, van Dijk, Steenbeek, & van Geert, 2014; Musz & Thompson-Schill, 2015; Steenbeek, van Geert, & van Dijk, 2011; Steenbeek, Jansen, & van Geert, 2012; Steenbeek & van Geert, 2013; Steenbeek, van der Aalsvoort , & van Geert, 2014) show that the variability of learning inputs affect that of learning outcomes. For instance, the variability of learning contents affects complexity levels of children's explanations on the contents (Meindertsma et al., 2014); the study investigates how much levels of learners' explanation change when learning topics vary, for instance, from a topic about a shape of an object to another topic about a weight of an object. Furthermore, another study shows that the variability of instruction types (e.g., open question, closed question, encouragement, etc.) influences complexity levels of task performance of learners (Meindertsma et al., 2013). There are other studies on the effect of the learning inputs on learning outcomes. For example, the variability of teachers' actions in a classroom has an influence on that of learners' actions (Steenbeek et al., 2011; Steenbeek et al., 2012; Steenbeek & van Geert, 2013; Steenbeek et al., 2014). Finally, it is found that spoken texts have variability that affects readers' emotions (Drożdż et al., 2016). For instance, a text with many technical words, which has low variability, can make readers bored. All of those studies mentioned above illustrate that the variability of learning inputs affects that of learning outputs.

Traditional vs. Present Views on Variability

Traditional educational studies used to regard variability as measurement error (van

Dijk & van Geert, 2007; van Dijk & van Geert, 2015). The main approaches in these studies are applying basic statistical techniques such as descriptive statistics, assuming that the average can truly capture the nature of change (van Geert & van Dijk, 2002).

Van Dijk and van Geert (2015) point out that the notion that variability is caused by measurement error originates from the true score theory (Cronbach, 1960; Lord & Novick, 1968). The essential argumentation of the true score theory is that every observed performance consists of a true performance and measurement error (van Dijk & van Geert, 2015). For example, a learner might have acquired a certain level of conceptual understanding, and if this is not manifest in a different measurement occasion, that is regarded as measurement error in the true score theory, even though conceptual understanding can vary in a specific measurement condition (i.e., the subject feels fatigued, the time frame of measurement is too long or too short, etc.).

One of the problems of the true score theory lies in the point that variability is regarded as just a random phenomenon or an unimportant piece of "extra" information by reducing variability to measurement error (van Dijk & van Geert, 2015). In reality, even measurement error may not be random but may be caused by specific factors (as mentioned above, the mental and physical conditions and the time frame of measurement are examples of such factors). Another problem is that the true score theory attempts to average out variability by applying data smoothing techniques, which causes the situation that the observed variability is neglected or at most regarded as a surface phenomenon. Even though observed behaviors vary, the underlying factors (e.g., latent psychological variables) of the variability can be more or less constant (van Dijk & van Geert, 2015).

In sum, traditional educational studies are likely to view variability as measurement error or unimportant information, which derives from a couple of the limitations of the true score theory that do not delve into the deep nature of variability.

Current views on variability are different from traditional ones in that the current views regard variability as essential information to explain the process of change and learning (Fischer & Bidell, 2006; Kunnen, & van Geert, 2012; van Dijk & van Geert, 2007; van Dijk & van Geert, 2015). Some educational researchers have progressively grasped the concept of variability and applied advanced methods of complexity science (Koopmans & Stamovlasis, 2016). For instance, standardized dispersion analysis

(Delignieres, Fortes, & Ninot, 2004; Delignieres, Ramdani, Lemoine, Torre, Fortes, & Ninot, 2006; Van Orden, Holden, & Turvey, 2003; Wijnants, Cox, Hasselman, Bosman, & Van Orden, 2012), recurrence quantification analysis (Angus, Smith, & Wiles, 2012; Boker, Xu, Rotondo, & King, 2002; Dale & Spivey, 2005; Marwan, Romano, Thiel, & Kurths, 2007; Trulla, Giuliani, Zbilut, & Webber, 1996; Wijnants, Hasselman, Cox, Bosman, & van Orden, 2012), detrended fluctuation analysis (Bryce & Sprague, 2012; Najafi & Darooneh, 2017; Shao, Gu, Jiang, Zhou, & Sornette, 2012), etc. All of these methods——sometimes called as *nonlinear dynamics methods*——are under the same umbrella of complexity science (Koopmans & Stamovlasis, 2016).

One of the essential aspects of these methods is not to neglect variability, but instead to treat it as valuable information to elucidate the process of change and learning (Kunnen, & van Geert, 2012; van Dijk & van Geert, 2015; van Geert & van Dijk, 2002). Whereas traditional approaches just applying basic statistical techniques cannot capture variability, those methods referred as nonlinear dynamics methods can do so. Not only from this aspect but also from another aspect, those methods can be considered to be suitable for this study. As discussed above, most of the previous studies on MOOCs that use traditional methodologies have difficulty in capturing the dynamics of lectures (Veletsianos & Shepherdson, 2016). One of the possible contributions of nonlinear dynamics methods is to overcome the difficulty from which previous research on MOOCs suffer and to expand the scope of research on MOOCs by examining the variability that can be seen in MOOC lectures.

Types of Variability

Variability can be classified into two categories. One is quantitative variability, and the other is qualitative variability (van Geert & van Dijk, 2002). The essence of quantitative variability is that it is one-dimensional. Each measurement consists of a degree of level on a single dimension. It can be a frequency count during an observation. For instance, the number of abstract concepts varies in each sentence when we examine a text. On the other hand, qualitative variability is multi-dimensional. The point is that each measurement consists of a set of different variables. For example, when we analyze a learner's behavior in class while he or she is solving a mathematical problem, the learner may use cognitive strategy A in 20% of measurement occasions, strategy B in 30%, and

strategy C in 50%. As van Geert and van Dijk (2002) argues, the most significant difference between the two variability categories is that qualitative variability deals with additional dimensions. In other words, it introduces new and different variables in measurement occasions; these variables can disappear and appear again during an observation. In terms of these definitions, the variability in our work is classified as one-dimensional quantitative variability because this study regards each variability of different measures (sentence length, stopwords, and domain-specific words) as a single dimension of measurement. Another reason is that one variable consisting of variability in this study does not disappear and reappear in measurement occasions; for example, it may be possible that the number of domain-specific words in a sentence is 0, but it does not mean that the variable itself disappears.

To further delve into the nature of variability, it can be categorized into three main types: (1) *white noise*, (2) *pink noise*, and (3) *brown noise* (Friedenberg, 2009; Hollis, Kloos, & van Orden, 2009; Mainzer, 2009; Ward, 2002). First, white noise has the highest fluctuation among the three. In other words, white noise is the most unstable variability pattern (e.g., the situation in which the frequency of unknown concepts in a sentence unsteadily increases and decreases over the course).

Second, pink noise is a typical feature of complex dynamical systems. In fact, it can be seen in various behaviors of complex dynamical systems such as the music of Bach (Mainzer, 2009), healthy heart beats (Goldberger, Amaral, Hausdorff, Ivanov, Peng, & Stanley, 2002), alpha brain waves (Ward, 2002), and spoken words (Angus et al., 2012; Drożdż et al., 2016). Pink noise can be conceived as optimal balance between stableness and unstableness of variability.

Third, brown noise has the lowest fluctuation among the three. To put it differently, it is the most stable variability pattern (i.e., the situation in which a lecturer repeatedly introduces the excessive amount of familiar concepts for learners over the course).

The following will discuss the effect of variability of three metrics used in this study: (1) sentence length, (2) stopwords, and (3) domain-specific words.

Sentence Length

Mikk (2008) found the effect of sentence length on working memory. The research shows that too long and too short sentences are detrimental to working memory (Mikk,

2008). However, this study does not examine the variability of sentence length. Instead, this study focuses only on the effect of long and short sentences.

Cognitive load theory (Sweller, Van Merriënboer, & Paas 1998) points out the influence of working memory on learning. In fact, some research reveals that the more the cognitive load in working memory increases the less learners achieve from learning (Diamond, 2013). Based on these findings, it can be hypothesized that sentence length has an effect on learning outcomes, but previous systematic reviews (Liyanagunawardena et al., 2013; Gasevic et al., 2014) show that the hypothesis in the context of MOOCs has not been tested so far.

Previous studies indicate the significant impact of variability in texts on learning. For instance, the variability of sentence length in texts has an effect on reading skills (Foorman, Francis, Davidson, Harm, & Griffin, 2004). Although Foorman et al. (2004) examined the characteristics of sentence in detail (e.g., lexical, semantic, and syntactic features) and investigated the effect of variability on reading achievement, their study lacks the investigation on variability types such as pink, white, and brown noise. For example, their study just examined the distribution of word frequency, which is not a deep structure of the variability of sentence length but a surface structure. As Fischer and Bidell (2006) point out, overlooking a deep structure of variability may cause a false conclusion on a cause and effect relationship when we examine the effect of variability on learning outcomes.

Musz and Thompson-Schill (2015) show a captivating finding that semantic variability in texts - that corresponds to sentence length - correlates with neural variability. Their study defines semantic variability as the variety of contexts in which each concept occurs. Musz and Thompson-Schill (2015) discovered that semantic variability in texts has an impact on neural activities in particular semantic memory, which affects conceptual understanding. In addition to this finding, their study was important in that it did not regard the variability as measurement error. Instead, their study considered the variability as a crucial factor to explain the effects on semantic memory. However, this study did not closely examine the types of variability.

Another research indicates that semantic variability in texts relates to semantic ambiguity that has a negative impact on semantic processing (Hoffman, Lambon Ralph,

& Rogers, 2013). Hoffman et al. (2013) defined semantic variability in a different way, compared with the definition of Musz and Thompson-Schill (2015). Hoffman et al. (2013) regarded semantic variability as the phenomenon that the meanings of words vary depending on contexts. Also, semantic variability is related to sentence length because it determines the nature of context (Hoffman et al., 2013); the more the sentence length is, the richer the meaning of the context can be (e.g., "This apple is red." versus "This apple that I purchased at a supermarket yesterday is red, whose color makes me feel invigorated.").

These previous studies find out that the variability in texts is inherently related to learning performance. However, these studies overlook the deep structure of variability, which is variability types. In other words, their studies reduce variability to one single concept. As discussed earlier, variability has diverse characters, depending on its types; e.g., white noise has high variability whereas brown noise has low variability.

Based on the findings and limitations of the past studies, this study postulates that the variability of sentence length in lecture videos might have an impact on learning outcomes.

Stopwords

Stopwords are the words that carry low information contents such as "a" or "the" (Makrehchi & Kamel, 2017; Duwairi, 2006). These words can hamper to construct a main topic in a text because they cause unrelated words to the topic (Xu, Y. Yin, J. Yin, 2017). Xu et al. (2017) point out that stopwords tend to occupy positions in a sentence, which leads to obscuring what the main topic is in the text. The list of stopwords in their study include not only typical stopwords such as "a" and "is" but also other types of stopwords such as "because" and "if," which can play an important role to construct a topic in a sentence.

Haas and Losee (1994) insist that strings of content words are more informative than those of stopwords. For instance, the string *the dynamics of social networks* contains two content words——*dynamics* and *social networks*——, and the string represents a specific concept. On the other hand, the string *it is there* contains no content words, and the string does not describe any particular topic. Haas and Losee's (1994) study indicates that interrupted information by the strings with too many stopwords made people have

difficulty in identifying and classifying a topic in a sentence; the list of stopwords in their study covers the same type of stopwords that Xu et al. (2017) utilized. Their study supports the claim of Xu et al. (2017) that common stopwords hinder a topic formation in a text.

In addition, Fraundorf and Watson (2011) argue that stopwords affect working memory. Too many stopwords in a sentence can be interruptions for learners to understand and recall a topic. The more stopwords appear in a discourse, the less learners remember the topic (Fraundorf & Watson, 2011). Because working memory has an impact on learning performance in online environments (Cevik & Altun, 2016) and because stopwords have an influence on not only working memory (Fraundorf & Watson, 2011) but also a topic formation (Xu et al., 2017), it would be plausible that the number of stopwords in a sentence have some effects on learning outcomes.

Also, all of the studies (Fraundorf & Watson, 2011; Haas & Losee, 1994; Xu et al., 2017) mentioned above examined just the number of stopwords in a sentence. In other words, they did not investigate a variability pattern that sequential sentences generate. As discussed earlier, their research methods can be categorized into a traditional statistical approach on learning (Fischer & Bidell, 2006) because they neglect variability. Neglecting variability of stopwords would be problematic because spoken words that inevitably contain stopwords create variability (Angus et al., 2012; Drożdż et al., 2016).

Domain-Specific Words

Auble and Franks (1983) found that the number of domain-specific words in a sentence had an influence on the degree of comprehension of a text. Learners comprehend a sentence when they have a schema that can account for the information in the sentence (Auble & Franks, 1983). The more the number of domain-specific words in a sentence increases, the more schemas learners have to possess. In terms of cognitive load theory (Sweller et al., 1998), too many domain-specific words in a sentence may harm cognitive loads of learners, which can impede learning because the more the cognitive load in working memory increases the less learners achieve from learning (Diamond, 2013). In fact, the number of domain-specific words in a sentence determines the degree of sentence comprehension for learners (Auble & Franks, 1983). Their study (Auble & Franks, 1983) suggests that providing too many domain-specific words in a

sentence should be avoided because learners take time to digest each domain-specific word and because semantic networks are gradually constructed.

As well as stopwords, domain-general words——which are broad concepts such as "evidence," "argumentation," "theory," "model," etc.——can impede a topic formation in a text because they are unrelated words to the topic (Xu et al., 2017). Xu et al. (2017) insist that not only stopwords but also domain-general words are likely to occupy positions in a sentence, which makes the main topic obscure. To put it differently, the amount of domain-specific words is essential for a topic formation, and if there are too many domain-general words in a sentence, understanding of the topic of the sentence can be inhibited.

As discussed above, previous research indicates that the number of domain-specific words in a sentence is key to the degree of sentence comprehension (Auble & Franks, 1983), and that the proportion of domain-specific words in a sentence affects a topic formation (Xu et al., 2017). Yet, these studies do not delve into the variability of domain-specific words in a sentence. As mentioned earlier, because the variability of learning inputs affect the variability of learning outcomes, it is crucial to scrutinize the variability of domain-specific words in a sentence when we analyze the effect on learning.

Hypothesis

As the above literature review shows, sentence length, stopwords, and domain-specific words can influence learning outcomes. This study hypothesizes that not only those variables but also their variability affects learning outcomes because the dynamic fluctuation of those variables may influence learners' cognitive load in a certain way as discussed above. In order to test the hypothesis, this study applies standardized dispersion analysis (Delignieres, et al., 2004; Delignieres, et al., 2006; Van Orden, et al., 2003; Wijnants, et al., 2012) that is a nonlinear dynamics method that can reveal the nature of variability.

Research Questions

The main purpose of this study is to examine the effects of the variability of sentence length, stopwords, and domain-specific words in the transcribed texts on learning outcomes. This study addresses the following three research questions.

1. What kind of variability type (e.g., white, pink, and brown) do the transcribed

texts have?

2. To what extend does the variability scale of sentence length, stopwords, and domain-specific words correlate with completion rates, quiz scores, and test scores?

3. What is the best predictor variable for completion rates, quiz scores, and test scores?

Method

Materials

This study examined the fourth run of the MOOC "Decision Making in a Complex and Uncertain World" offered by University of Groningen from 2016 to 2017; it was the latest version at the beginning of this study. The MOOC lasted seven weeks, and each week consisted of a series of steps such as lecture videos, articles, discussions, quizzes, and a test. The main learning goal of the course was to teach important concepts of complexity science and how to make a decision in a complex world.

This study investigated transcribed texts of lecture videos in all seven weeks. Lecture videos were either content videos or outline videos. The former is the videos that are relevant to topics in each week, whereas the latter is the videos that offer information about logistics of the course and provide a brief introduction or summary of the contents in each week. Because this study aims at examining the effects of the contents in lectures, only content videos were included in analysis.

Participants

The total number of registrants for the course were 8655, whereas the total number of active learners were 2594 who have completed at least one step within 180 minutes at anytime in any course work. The number of participants who completed more than 50% of the course materials were 498, and that of those who completed more than 90% of the course materials was 313. Because the course was offered to people all over the world, their age, nationality, and educational level were diverse; the age range of the largest proportion of the participants was between 26 and 35, half of the participants came from Europe, and the highest educational level of the largest proportion of the participants was a bachelor's degree.

Procedure

We obtained the transcribed data from the website of the MOOC. Outcome measures were provided by the department of Educational Support and Innovation at University of

Groningen

All of the time-series data were generated from the transcripts of lecture videos. The transcripts were public information; any learners could see them on the webpage of each lecture video. R, a software for statistical computing, quantified each sentence in all of the lecture videos in terms of sentence length, stop-words, and domain-specific words. Standardized dispersion analysis and statistical analysis were conducted with R.

Measures

This study has three dependent variables as outcome measures: (1) completion rates, (2) quiz scores, and (3) test scores; and three independent variables: (1) sentence length, (2) stopwords, and (3) domain-specific words.

Completion rate

Learners could push a "mark as complete" button on the screen once they watched a lecture video. Completion rate was calculated by dividing the number of learners that marked a video as complete by the total number of learners who started to view the video. This study only included the completion rates of content videos. The outcome measure of completion rate is the average ratio in each week.

Quiz score

Each week had a couple of quiz questions between lecture videos. Quiz questions were formative assessments so that learners could test their understandings of the contents of lectures. There was no restriction on the number of attempts for a learner to take a quiz. The outcome measure of quiz score derived from only the first attempt. The average quiz score in each week was calculated.

Test score

At the end of each week, a learner could take a test. Test questions were summative assessments so that learners could test their understandings of what they learned through the contents of the entire week. The maximum number of attempts to take a test was three. The outcome measure of test score was generated from only the first attempt. The average test score in each week was calculated.

Sentence length

Sentence length is a measure of the number of words in a sentence. This study used an R function to count the number of words in each sentence.

Stopwords

R has a package called "stopwords," which consists of a comprehensive list of English stopwords (See Appendix A). The number of stopwords in each sentence was counted by an R function.

Domain-specific words

This study defines domain-specific words as key words in the academic field that the MOOC focuses on. Although this course addresses how to make a decision in a complex world, it is not the main topic. The course mainly focuses on theoretical aspects of complexity science rather than practical ones. In fact, the course offers ample theoretical knowledge about complexity science. For instance, the actual course contents cover the following topics; "complex system (week 1)," "emergent behavior (week 2)," "agent-based modeling (week 3)," "evolutionary dynamics (week 4)," "cellular automata (week 5)," "path dependence (week 6)," "nonlinearity (week 7)."

Since the chief aim of this course is to provide knowledge of complexity science, domain-specific words extracted from key terminologies in the field. The list of domain-specific words was generated from Goldstein's (2008) literature study and the glossary list provided by the Santa Fe Institute (2018; https://www.complexityexplorer.org/explore/glossary), which is one of the most renowned research centers in the field of complexity science. Duplicated words were eliminated, and the total number of domain-specific words was 218 (See Appendix B).

Analytical Technique

Standardized dispersion analysis

Standardized dispersion analysis (SDA) evaluates a scale of variability within a time-series (Delignieres et al., 2006; Van Orden et al., 2003; Wijnants et al., 2012). Variability is measured by the mean and standard deviation of time-series. The first step is to normalize the time-series with the mean and standard deviation. The second step is to calculate each mean between two consecutive data points. For instance, if the time-series has 512 data points, this step is to calculate a mean 256 times. Then, the standard deviation of 256 mean values is computed. The following procedure is the same; each mean between four, eight, sixteen, thirty-two, etc., and two hundred fifty-six consecutive data points is calculated, and each standard deviation is computed. The final step is to

calculate the base 10 logarithm of each standard deviation. The following figure is an example of log-log plot to measure a scale of variability.

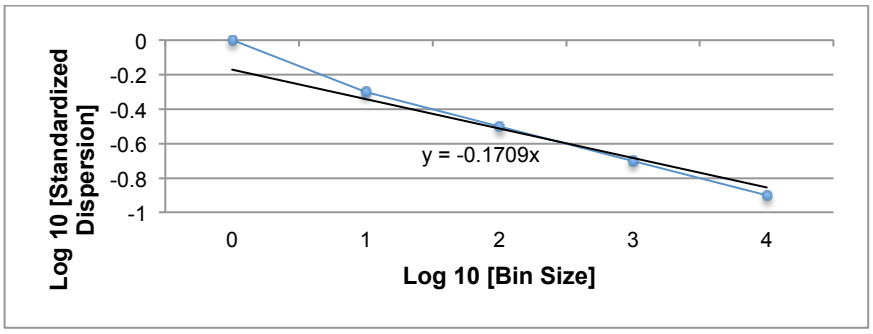


Figure 1. Example of standardized dispersion analysis

The x-axis in Figure 1 indicates the base 10 logarithm of the number of data points of each consecutive mean. The y-axis illustrates the base 10 logarithm of the standardized dispersion for each bin size. The black line indicates a least squares regression line for five points represented by the blue line. The slope of the black line is called "spectral slope," which is the variability scale of the time-series (Wijnants et al., 2012). This study used the package "fractal" in R to calculate the variability scale of each content lecture video and then computed the average variability scale in each week.

Variability scale

Variability scales have three types corresponding to their values (Van Orden et al., 2003; Wijnants et al., 2012). The range of the scale is from -2 to 0. A scale close to -2 means that the time-series varies randomly, which is called "brown noise." A scale close to -1 is called "pink noise" or "1/f noise." Lastly, a scale is close to 0 is called "white noise," which represents a highly fluctuating but not random time-series.

Results

Before addressing the first research question, we examined descriptives of lecture videos in each week.

Table 1

Descriptives of Lecture Videos in Each Week

Week	Number of content videos	Number of sentences	Average SL in a sentence	Average SW in a sentence	Average DSW in a sentence
1	8	353	13.75	8.65	0.24
2	11	439	12.52	7.87	0.42
3	15	361	14.66	9.40	0.23
4	6	415	17.86	10.36	0.26

5	8	375	15.75	9.35	0.26
6	13	502	17.11	9.85	0.11
7	9	475	12.37	7.78	0.15

Note. SL = sentence length; SW = stopwords; DSW = domain-specific words.

Table 1 shows the total number of content videos and sentences in each week. Also, Table 1 describes each week's average number of sentence length, stopwords, and domain-specific words in a sentence. The following correlation analysis was conducted based on these average metrics.

Table 2

Correlation Coefficient between Each Metric

Variable	Correlation coefficient	p-value	95% CI
Average SL & average SW	0.97	**.0002**	[0.83, 0.99]
Average SL & average DSW	-0.32	.48	[-0.86, 0.57]
Average SW & average DSW	-0.30	.51	[-0.86, 0.59]

Note. SL = sentence length; SW = stopwords; DSW = domain-specific words.

The results show that sentence length and stopwords had a strong and significant positive correlation. On the other hand, no significant correlation was found between sentence length and domain-specific words, and between stopwords and domain-specific words.

A sequence of sentences is treated as a time-series in this study. The following Figure 2 shows an example of the time-series of sentence length in the first week lectures, which consists of eight lectures and 353 sentences in total. All of the time-series data are reported in Appendix C.

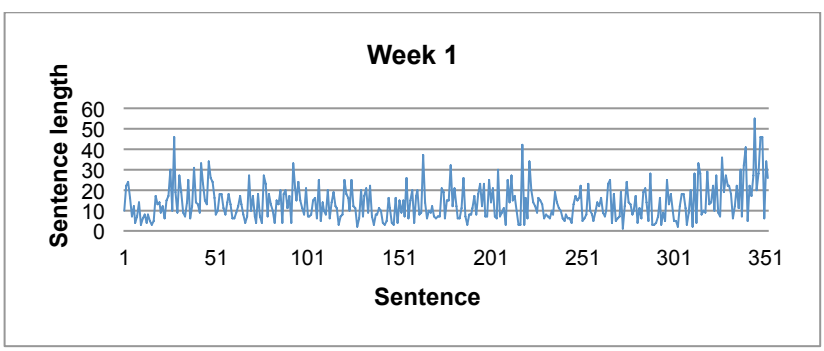


Figure 2. Time-series data of sentence length in the first week

Figure 2 shows that sentence length in the time-series data dynamically varies over the first week. For instance, some sentences have short sentence length, whereas others

have long sentence length. Although Figure 2 tells us that the time-series data have variability, the degree of the variability is still unknown when we just look at the time-series data.

The first research question was: "What kind of variability type (e.g., white, pink, and brown) do the transcribed texts have?" The following table indicates learning outcomes and the results of the variability scales of each metric.

Table 3

Descriptives of learning outcomes and variability scores for each week

Week	Completion rate	Quiz score	Test score	Variability scale of SL	Variability scale of SW	Variability scale of DSW
1	86.75	78.68	88.53	-0.74	-0.64	-0.65
2	92.09	80.70	89.85	-0.55	-0.62	-0.33
3	93.27	81.12	81.85	-0.72	-0.69	-0.45
4	92.17	86.27	82.93	-0.74	-0.78	-0.39
5	92.63	74.85	89.83	-0.46	-0.55	-0.36
6	90.77	80.03	87.48	-0.62	-0.62	-0.45
7	92.56	74.86	NA	-0.67	-0.58	-0.46

Note. SL = sentence length; SW = stopwords; DSW = domain-specific words.

Table 3 shows that there is no brown noise in the entire weeks because all of the three variability scales in each week are not close to -2. To put it differently, the underlying patterns of weekly lectures in terms of three metrics were not random but had a certain pattern of pink noise or white noise. More closely examined, the variability scale of sentence length revealed that most of the weeks except for the fifth week indicated pink noise because the values were less than -0.5——closer to -1. The variability scale of stopwords showed that all of the weeks had pink noise because the values were less than -0.5. On the other hand, the variability scale of domain-specific words showed that most of the weeks except for the first week had white noise because the values were more than -0.5——closer to zero.

Before we address the second research question, we need to know whether the variability scales correlate. The following table shows the correlation between all of the three variability scales.

Table 4

Correlation Coefficient between Variability Scales

Variable	Correlation coefficient	p-value	95% CI
Variability scale of SL & variability scale of SW	0.71	.07	[-0.09, 0.95]

Variability scale of SL & variability scale of DSW	0.62	.13	[-0.25, 0.94]
Variability scale of SW & variability scale of DSW	0.03	.94	[-0.74, 0.77]

Note. SL = sentence length; SW = stopwords; DSW = domain-specific words.

The results showed no significant correlation between the three variability scales. Then we can address the second research question: "To what extend does the variability scale of sentence length, stopwords, and domain-specific words correlate with completion rates, quiz scores, and test scores?"

Table 5

Correlation Coefficient between Learning Outcomes and Variability Scales

Variable	Completion rate	Quiz score	Test score
Average SL	0.08	0.52	-0.43
Average SW	0.09	0.57	-0.62
Average DSW	0.09	0.25	0.29
Variability scale of SL	0.34	-0.52	0.71
Variability scale of SW	-0.02	**-0.94****	**0.83***
Variability scale of DSW	**0.83***	0.14	0.03

Note. SL = sentence length; SW = stopwords; DSW = domain-specific words. *p < .05, **p < .01.

The results showed that the average number of all of the metrics did not have any significant correlation with learning outcomes. However, the variability scale of domain-specific words had a strong and significant positive correlation with the completion rate; 95% confidence interval [CI] -0.21 to 0.97, p = .02, t = 3.33, df = 5. In addition, the variability scale of stopwords had a strong and significant negative correlation with the quiz score; CI -0.99 to -0.63, p = .002, t = -6.03, df = 5. Lastly, the variability scale of stopwords had a strong and significant positive correlation with the test score; 95% confidence interval [CI] 0.05 to 0.98, p = .04, t = 2.97, df = 4.

The next analysis addressed the last research question: "What is the best predictor variable for completion rates, quiz scores, and test scores?"

Table 6

Simple Regression Analysis

	Completion rate			Quiz score			Test score		
	R^2	B	β	R^2	B	β	R^2	B	β
SL	.11	6.98	.34	.27	-19.13	-.52	.50	21.46	.71
SW	.00	-0.57	-.02	.88	**-48.88****	-.94	.69	**37.28***	.83
DSW	.69	**17.57***	.83	.02	5.21	.14	.00	0.99	.03

Note. SL = variability scale of sentence length; SW = variability scale of stopwords; DSW = variability scale of domain-specific words. *p < .05, **p < .01.

The results of the simple regression analysis showed that the variability scale of

stopwords significantly predicted the quiz score (β = -.94, p = .002) and the test score (β = .83, p = .04). The variability scale of stopwords also explained a significant proportion of the variance in the quiz score (R= .88, F(1, 5) = 36.41, p = .002) and the test score (R= .69, F(1, 4) = 8.82, p = .04). Furthermore, the variability scale of domain-specific words significantly predicted the completion rate (β = .83, p = .02). The variability scale of domain-specific words also explained a significant proportion of the variance in the completion rate (R= .69, F(1, 5) = 11.11, p = .02). However, the variability scale of sentence length did not significantly predict any learning outcomes.

Finally, a multiple regression analysis was conducted in order to discover models to predict learning outcomes.

Table 7

Multiple Regression Analysis

Model 1. SL & SW									
	Completion rate			Quiz score			Test score		
	ΔR^2	B	β	ΔR^2	B	β	ΔR^2	B	β
Step1	.11			.27			.50		
Constant		95.96***			67.20***			100.45***	
SL		6.98	.34		-19.13			21.46	.71
Step2	.14			.66			.19		
Constant		91.14***			48.25***			110.47***	
SL		14.78	.71		11.55	.31		3.55	.12
SW		-15.35	-.53		-60.42**	-1.16		33.01	.73
p-value	.56			**.005****			.17		
R^2	.25			.93			.69		

Model 2. SL & DSW									
	Completion rate			Quiz score			Test score		
	ΔR^2	B	β	ΔR^2	B	β	ΔR^2	B	β
Step1	.11			.27			.50		
Constant		95.96***			67.20***			100.45***	
SL		6.98	.34		-19.13			21.46	.71
Step2	.63			.33			.27		
Constant		97.02***			68.61***			99.43***	
SL		-6.03	-.29		-36.22	-.98		33.68*	1.11
DSW		21.38*	1.01		28.07	.74		-20.11	-.66
p-value	.07			.16			.11		
R^2	.74			.60			.77		

Model 3. SL, SW, & DSW									
	Completion rate			Quiz score			Test score		
	ΔR^2	B	β	ΔR^2	B	β	ΔR^2	B	β
Step1	.11			.27			.50		
Constant		95.96***			67.20***			100.45***	
SL		6.98	.34		-19.13			21.46	.71

Step2	.14			.66			.19		
Constant		91.14***			48.25***			110.47***	
SL		14.78	.71		11.55	.31		3.55	.12
SW		-15.35	-.53		-60.42**	-1.16		33.01	.73
Step3	.60			.00			.10		
Constant		104.02***			47.20***			89.53	
SL		-22.63	-1.09		14.61	.39		58.86	1.95
SW		20.70	.71		-63.37*	-1.22		-30.21	-.67
DSW		31.39*	1.48		-2.57	-.07		-34.58	-1.13
p-value	.09			**.03***			.30		
R^2	.85			.93			.79		

Model 4. SW & DSW

	Completion rate			Quiz score			Test score		
	ΔR^2	B	β	ΔR^2	B	β	ΔR^2	B	β
Step1	.00			.88			.69		
Constant		91.10***			48.22***			110.98***	
SW		-0.57	-.02		-48.88**	-.94		37.28*	.83
Step2	.69			.02			.00		
Constant		98.37***			50.84***			110.74**	
SW		-1.33	-.05		-49.15**	-.94		37.34	.83
DSW		17.60*	.83		6.33	.17		-0.62	-.02
p-value	.10			**.009****			.17		
R^2	.69			.91			.69		

Note. SL = variability scale of sentence length; SW = variability scale of stopwords; DSW = variability scale of domain-specific words. *p < .05, **p < .01, ***p < .001

The results showed that there were no significant models to predict the completion rate and the test score. On the other hand, there were three models to significantly predict the quiz score. The first one was the combination of sentence length and stopwords ($F(2, 4) = 25.34$, $p = .005$), and the model explained 93% of the variance. While sentence length did not significantly contribute to the regression model, stopwords did ($\beta = -1.16$, $p = .004$). The second one was the combination of sentence length, stopwords, and domain-specific words ($F(3, 3) = 12.91$, $p = .03$), and the model explained 93% of the variance. Only stopwords significantly contributed to the model ($\beta = -1.22$, $p = .03$). The last one was the combination of stopwords and domain-specific words ($F(2, 4) = 19.56$, $p = .009$), and the model explained 91% of the variance. Whereas stopwords significantly contributed to the model ($\beta = -.94$, $p = .003$), domain-specific words did not. In the three types of combination, only stopwords significantly contributed to the regression models.

In sum, the predictive models for the quiz score were:

1. Quiz score = (.31 * Sentence length) + (-1.16 * Stopwords)
2. Quiz score = (.39 * Sentence length) + (-1.22 * Stopwords) + (-.07 * Domain-specific words)
3. Quiz score = (-.94 * Stopwords) + (.17 * Domain-specific words)

Discussion

First, a strong and significant positive correlation was found between sentence length and stopwords. It implies that the number of stopwords in a sentence increases when that of sentence length increases. This finding is important because previous studies on sentence lengh and stopwords (i.e., Fraundorf & Watson, 2011; Mikk, 2008; Xu et al., 2017) did not examine the relationship. On the other hand, there was no significant correlation between sentence length and domain-specific words and between stopwords and domain-specific words. This result indicates that domain-specific words in the lecture videos in the MOOC were independent of sentence length and stopwords. In other words, the result implies that the number of stopwords does not necessarily increase when that of sentence length increases or that of stopwords decrease. If the number of stopwords decreases in a sentence, there is room for words other than stopwords to occur in the sentence. However, domain-specific words were not candidates to emerge in the sentence because there was no significant correlation between stopwords and domain-specific words. Words with semantically different functions such as domain-general words might have occurred in this case. This finding also contributed to shedding light on the relationship between the three measures, which had not been investigated by previous studies on domain-specific words (e.g., Auble & Franks, 1983; Xu et al., 2017).

Second, the correlation analysis on the variability scale of each metric showed that there was no significant correlation between each metric. Because the correlation between sentence length and domain-specific words and between stopwords and domain-specific words in terms of the average number in each week did not have any significant correlation, it is plausible that the correlation between the variables in terms of the variability scale may not also have any significant correlation. In fact, there was no correlation between the variability scales. However, if there were significant correlation between the variables in terms of the average number of each metric, it would be likely that there is significant correlation between the variability scales of the variables. Yet, it was counter-intuitive that no correlation was found between the variability scale of sentence length and that of stopwords even though there was a strong and significant positive correlation between the average number of sentence length and that of stopwords. This result implies that the variability scale had a unique feature in that it

cannot be reduced to the original time-series. In fact, it is the distinctive nature of time-series, which is called a fractal dimension (Delignieres at el., 2006; Kelso, 1995; Mainzer, 2007; Van Orden et al., 2003; Ward, 2002; Wijnants et al., 2012). One of the key implications here is that it is important to examine a fractal dimension of time-series in order to reveal a unique relationship between variables that cannot be discovered when we look at a surface character of time-series.

The analysis on the types of variability scale revealed that each metric had a unique fractal dimension. Although there were studies on the impact of variability in texts on learning (e.g., Foorman et al., 2004; Hoffman et al., 2013; Musz & Thompson-Schill, 2015), these studies did not closely examine variability types. Instead, they treated variability as one single entity. Almost all of the weeks except for the week five in terms of the variability scale of sentence length showed pink noise——which shows an optimal balance between stableness and unstableness of variability——, and all of the weeks for the variability scales of stopwords also indicated pink noise. On the other hand, almost all of the weeks other than the week first regarding the variability scale of domain-specific words showed white noise that is the most unstable variability pattern. As discussed above, both time-series of sentence length and stopwords correlated with each other at a surface level, but both fractal dimensions did not have any significant correlation. It suggests that not only the fractal dimension of domain-specific words but also that of sentence length and stopwords had a unique characteristic that cannot be reduced to each other. To summarize, the variability scale of each metric was independent and unique, which cannot be revealed when we examine the surface characteristic of each time-series. In this sense, the finding revealed the benefit of using SDA, and this method would be a new analytical tool to solve the issue of the methodological limitation that previous MOOC studies had (Veletsianos & Shepherdson, 2016).

The correlation analysis on the relationship between learning outcomes and the variability scale of each metric showed that the variability scale of domain-specific words had a strong and significant positive correlation with the completion rate. Because the variability scale in each week had white noise, the result indicates that if it became more white noise——in other words, the spectral slope became closer to zero——, the completion rate would increase. In fact, as the simple regression analysis showed, when

the variability scale of domain-specific words increases by one unit, the completion rate increases by 0.83 unit (β= .83). Considering the result of no significant correlation between the average number of domain-specific words and the completion rate, this result suggests the importance of the investigation on the variability scale of the variable to predict the completion rate. To put it differently, we cannot predict the completion rate when we examine the surface character of the variable——just the average number of domain-specific words——, but we can when we investigate the deeper character of the variable——the variability scale.

Another finding was that the variability scale of stopwords had a strong and significant negative correlation with the quiz score while the variability scale of stopwords had a strong and significant positive correlation with the test score. As the simple regression analysis discovered, when the variability scale of stopwords increases by one unit, the quiz score decreases by 0.94 unit (β= -.94), whereas when the variability scale of stopwords increases by one unit, the test score increases by 0.83 unit (β= .83). In sum, the variability scale had two opposite directions of correlation with quiz and test score and predicted both learning outcomes in a contrary direction. As well as the finding of the relationship between domain-specific words and the completion rate, we would not be able to predict the quiz and test score if we neglected the variability scale of stopwords.

Finally, multiple regression analysis showed that there were no significant models to predict the completion rate and the test score. On the other hand, there were three models to significantly predict the quiz score. In the three models, only the variability scale of stopwords significantly contributed to the models. It makes sense because, as the simple regression analysis showed, only the variability scale of stopwords was a significant variable to predict the quiz score.

The development of explanations of the impacts of variability on outcome measures identified in this study is beyond the scope of this thesis. A perspective of cognitive psychology on text comprehension (e.g., Kintsch & Van Dyke, 1978) is particularly promising because it can be assumed that variability affects information processing, thus making capacities (and limits) of working memory relevant. Further analyses on which channel learners preferentially use (spoken/audio or written) may be able to shed light on

the identified effects, because written texts rather than audio enable learners to more easily repeat the information. The differential effects of the features on the quiz and the test are relevant in this respect; students complete quizzes after each lecture (thus have less time available for study), while they complete tests after several lectures. It can be speculated that variability affects the construction of understanding, and specifically the integration of new information with knowledge. The results of this study may therefore be able to contribute to extant models of text comprehension.

Limitations

The first limitation is that the item difficulty within a quiz and a test in the MOOC was different. For instance, the item difficulty of quiz questions for a particular lecture was not always the same as that of questions for another lecture; this is also true to a test in each week. This situation might cause a threat to "statistical conclusion validity (Shadish, Cook, & Campbell, 2002)." More specifically, it might have unreliably strengthened or weakened the relationship between variables. In the case of this study, the correlation between stopwords and the quiz score, and that between stopwords and the test score might have been overly increased or decreased because of the different item difficulty.

The second limitation is about completion rates. As introduced in the section of measures, learners could push a "mark as complete" button on the screen after they watched a lecture video. Technically speaking, they could push it without watching a video. We can conceive two scenarios: (1) Some learners might have pushed the button even if they did not watch a video, and (2) Some learners might have forgotten to push it even though they watched a video. We did not know how many learners did these actions.

The third limitation is related to construct validity and measurement criteria. Sentence length is a clear construct, and the measurement criterion is robust. On the other hand, the construct and measurement criterion of stopwords are not so clear compared to those of sentence length. Especially, the number of English stopwords is slightly different in various existing lists; this study utilized the list in R. In addition, as well as the previous studies (Xu et al., 2017; Haas & Losee, 1994), this study did not separate stopwords in terms of semantic functions in a sentence. For example, a stopword, "a"

may not help to form a topic in a sentence, but "if" and "because" may do so. It might be plausible that the variability of the former type of stopwords and that of the latter type have a distinct effect on learning outcomes. In sum, future study can thoroughly examine the effect of the variability of stopwords by separating them in terms of semantic functions.

Not only stopwords but also domain-specific words have neither a single definition nor an explicit criterion for a measurement. As the name of domain-specific words implies, a list of domain-specific words varies depending on a domain. The intrinsic variation might have lowered the construct validity. Furthermore, the list of domain-specific words (See Appendix B) might not have been comprehensive even though it was extracted from previous research (Goldstein, 2008) and the glossary generated by one of the renowned research institutes in the targeted domain for this study. For instance, several terms (e.g., risk, probability, unpredictability, etc.) with some important meanings were repeatedly discussed in the course, but these terms were not included in this study because the previous research and glossary did not cover these terms.

In addition, this study might have involved "construct confounding (Shadish et al., 2002)." In the context of this research, the meaning of construct confounding is that the list of domain-specific words is not a pure representation of constructs. Several words such as "information" and "interaction" (See Appendix B) may have resulted in incomplete construct representations because these words can be used not only in a specific domain but also in various domains. The construct confounding would be another threat to construct validity in this study.

To enhance the construct validity and to mitigate the construct confounding in this study, rigorous reclassification of domain-specific words is one solution. For instance, observation in the field on which a MOOC focuses and consultation with an expert in the domain would be helpful to thoroughly examine words after removing stopwords and to narrow down the words (e.g., lecture-specific words, course-specific words, and domain-specific words).

The final limitation is related to a methodological issue of SDA. One of the central issues to apply any nonlinear dynamics methods for evaluating fractal scales is the amount of data points (Delignieres et al., 2006; Bassingthwaighte & Raymond, 1995). To

put it differently, analyzing fractal scales requires a sufficient amount of data points, but the amount does not have clear cut-off criteria. The analytical package "fractal" in R requires at least 32 data points to apply SDA for calculating a variability scale. Some lectures did not have 32 data points (See Appendix D), and thus, their variability scales could not be measured. To overcome this limitation, Rangarajan and Ding (2000) suggest an integrated approach to apply another method to evaluate a fractal dimension in order to verify the accuracy. For instance, detrended-fluctuation analysis (Kirchner, Schubert, Liebherr, Haas, & Hernandez-Lemus; 2014; Najafi & Darooneh, 2017) would be one of the best candidates to apply an integrated approach (Rangarajan & Ding, 2000).

Summary and Conclusion

As introduced earlier, few studies have examined the effects of the formal features of MOOC lectures on learning outcomes. In addition, past research on MOOCs had a quite limited number of methodological approaches and a limited scope of research (Veletsianos & Shepherdson, 2016). This study intended to fill the gap and to propose a new approach to conduct a study on MOOCs.

Particularly, this study addressed the gap by examining the variability of transcribed spoken lectures. The results showed the significant effects of the variability of sentence length, stopwords, and domain-specific words on learning outcomes assessed by completion rates, quiz scores, and test scores.

This study applied standardized dispersion analysis (Delignieres et al., 2006; Van Orden et al., 2003; Wijnants et al., 2012) to examine a scale of variability, and this study showed that this method would be useful to investigate the effects of the variability of transcribed texts of MOOC lectures. To conclude, we hope that this study contributed to exploring the variability in MOOC lectures and proposing the possibility to use standardized dispersion analysis for examining variability in the context of MOOC research.

References

Angus, D., Smith, A., & Wiles, J. (2012). Conceptual recurrence plots: Revealing patterns in human discourse. *IEEE Transactions on Visualization and Computer Graphics, 18*(6), 988-997. doi:10.1109/TVCG.2011.100

Auble, P., & Franks, J. (1983). Sentence comprehension processes. *Journal of Verbal Learning and Verbal Behavior, 22*(4), 395-405. doi:10.1016/S0022-5371(83)90255-4

Bassingthwaighte, J., & Raymond, G. (1995). Evaluation of the dispersional analysis method for fractal time series. *Annals of Biomedical Engineering, 23*(4), 491-505. doi:10.1007/BF02584449

Bryce, R., & Sprague, K. (2012). Revisiting detrended fluctuation analysis. *Scientific Reports, 2*(1). doi:10.1038/srep00315

Cevik, V., & Altun, A. (2016). Roles of working memory performance and instructional strategy in complex cognitive task performance. *Journal of Computer Assisted Learning, 32*(6), 594-606. doi:10.1111/jcal.12156

Dale, R., & Spivey, M. (2005). Categorical recurrence analysis of child language. In *Proceedings of the Cognitive Science Society* (pp. 530–535).

Delignieres, D., Fortes, M., & Ninot, G. (2004). The fractal dynamics of self-esteem and physical self. *Nonlinear Dynamics, Psychology, and Life Sciences, 8*(4), 479-510.

Delignieres, D., Ramdani, S., Lemoine, L., Torre, K., Fortes, M., & Ninot, G. (2006). Fractal analyses for 'short' time series: A re-assessment of classical methods. *Journal of Mathematical Psychology, 50*(6), 525-544. doi:10.1016/j.jmp.2006.07.004

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135-68. doi:10.1146/annurev-psych-113011-143750

Drożdż, S., Oświęcimka, P., Kulig, A., Kwapień, J., Bazarnik, K., Grabska-Gradzińska, I., & Stanuszek, M. (2016). Quantifying origin and character of long-range correlations in narrative texts. *Information Sciences,* 331, 32-44. doi:10.1016/j.ins.2015.10.023

Duwairi, R, M. (2006). A framework for the computerized assessment of university student essays. *Computers in Human Behavior, 22*(3), 381-388. doi:10.1016 /j.chb.2004.09.006

Fischer, K. W., & Bidell, T. R. (2006). Dynamic development of action, thought, and emotion. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Theoretical models of human development* (6th ed, Vol.1, pp. 313-399). New York: Wiley.

Foorman, B. R., Francis, D. J., Davidson, K. C., Harm, M. W., & Griffin, J. (2004). Variability in Text Features in Six Grade 1 Basal Reading Programs. *Scientific Studies of Reading, 8(2)*, 167-197.

Fraundorf, S., & Watson, D. (2011). The disfluent discourse: Effects of filled pauses on recall. *Journal of Memory and Language, 65*(2), 161-175. doi:10.1016/j.jml.2011.03.004

Friedenberg, J. (2009). *Dynamical psychology: Complexity, self-organization and mind.* Litchfield Park, AZ: ISCE Publishing.

Gasevic, D., Kovanovic, V., Joksimovic, S., & Siemens, G. (2014). Where Is Research on Massive Open Online Courses Headed? A Data Analysis of the MOOC Research Initiative. *The International Review of Research In Open And Distance Learning, 15*(5), 134-176.

Glattre, H. R, & Glattre, E. (2018). Finding fractal networks in literature. Nonlinear Dynamics, Psychology, and Life Sciences, 22(2), 263-282.

Goldberger, A. L., Amaral, L. A. N., Hausdorff, J. M., Ivanov, P. Ch., Peng, C.-K., & Stanley, H. E. (2002). Fractal dynamics in physiology: alterations with disease and aging. *Proc. Natl. Acad. Sci. U.S.A. 99*, 2466–2472.

Goldstein, J. (2008). Resource guide and glossary for nonlinear/complex systems terms. In C. Lindberg, S. Nash, & C. Lindberg (Eds.), *On the edge: Nursing in the age of complexity* (pp. 263-294). Bordentown, NJ: Plexus Press.

González-Espinoza, A., Larralde, H., Martínez-Mekler, G., & Müller, M. (2017). Multiple scaling behaviour and nonlinear traits in music scores. *Royal Society Open Science, 4*(12), 171282-171282. doi:10.1098/rsos.171282

Haas, S., & Losee, R. (1994). Looking in text windows: Their size and composition. *Information Processing and Management, 30*(5), 619-619.

Hoffman, P., Lambon Ralph, M., & Rogers, T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods, 45*(3), 718-730. doi:10.3758/s13428-012-0278-x

Hollis, G., Kloos, H., van Orden, G.C. (2009). Origins of Order in Cognitive Activity. In S. Guastello, M. Koopmans, & D. Pincus (Eds.), *Chaos and complexity in psychology: The theory of nonlinear dynamical systems* (pp. 206-241). New York, NY: Cambridge University Press.

Kelso, S. J. A. (1995). *Dynamic patterns; The self-organization of brain and behavior.* Cambridge, MA: The MIT Press.

Kennedy, J. (2014). Characteristics of massive open online courses (MOOCS): A research review, 2009-2012. *Journal of Interactive Online Learning, 13*(1), 1-16.

Kintsch, W. & van Dijk, T. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363 – 394. doi:10.1037//0033-295X.85.5.363

Kirchner, M., Schubert, P., Liebherr, M., Haas, C., & Hernandez-Lemus, E. (2014). Detrended fluctuation analysis and adaptive fractal analysis of stride time data in parkinson's disease: Stitching together short gait trials. *Plos One, 9*(1), 85787. doi:10.1371/journal.pone.0085787

Koopmans, M., & Stamovlasis, D. (Eds.). (2016). *Complex dynamical systems in education: Concepts, methods, and applications.* Springer International Publishing.

Kunnen, S., & van Geert, P. (2012). General characteristics of a dynamic systems approach. In S. Kunnen (Ed.), *A dynamic systems approach to adolescent development* (pp. 15-34). East Sussex, UK: Psychology Press.

Liyanagunawardena, T. R., Adams, A., & Williams, S. A. (2013). MOOCs: A systematic study of the published literature 2008-2012. *The International Review of Research in Open and Distance Learning, 14*(3), 202-227.

Mainzer, K. (2007). *Thinking in complexity: The computational dynamics of matter, mind, and mankind (5th ed)*. New York: Springer.

Makrehchi, M., & Kamel, M. (2017). Extracting domain-specific stopwords for text classifiers. *Intelligent Data Analysis, 21*(1), 39-62. doi:10.3233/IDA-150390

Marwan, N., Carmen Romano, M., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports, 438*(5), 237-329. doi:10.1016/j.physrep.2006.11.001

Meindertsma, H. B., van Dijk, M. W. G., Steenbeek, H. W., & van Geert, P. L. C. (2013). Assessment of preschooler's scientific reasoning in adult-child interactions: What is the optimal context? *Research in Science Education, 44*, 215–237. doi:10.1007/s11165-013-9380-z

Meindertsma, H., van Dijk, M., Steenbeek, H., & van Geert, P. (2014). Stability and variability in young children's understanding of floating and sinking during one single-Task session. *Mind, Brain, and Education, 8*(3), 149-158.

Mikk, J. (2008). Sentence length for revealing the cognitive load reversal effect in text comprehension, *Educational Studies, 34*(2), 119-127. DOI: 10.1080/03055690701811164

Musz, E., & Thompson-Schill, S. L. (2015). Semantic variability predicts neural variability of object concepts. *Neuropsychologia,* 7641-51. doi:10.1016/j.neuropsychologia.2014.11.029

Najafi, E., & H. Darooneh, A. (2017). Long range dependence in texts: A method for quantifying coherence of text. *Knowledge-Based Systems, 133*, 33-42. doi:10.1016/j.knosys.2017.06.032

Rangarajan, G., & Ding, M. (2000). Integrated approach to the assessment of long range correlation in time series data. *Physical Review -Series E-, 61*, 4991-5001.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton Mifflin.

Shao, Y., Gu, G., Jiang, Z., Zhou, W., & Sornette, D. (2012). Comparing the performance of FA, DFA and DMA using different synthetic long-range correlated time series. *Scientific Reports, 2*(1). doi:10.1038/srep00835

Steenbeek, H. W., & Van Geert, P. L. C. (2005). A dynamic systems model of dyadic interaction during play of two children. *European Journal of Developmental Psychology, 2*(2), 105-145. DOI: 10.1080/17405620544000020

Steenbeek, H. W., & van Geert, P. L. C. (2007a). A theory and dynamic model of dyadic interaction: Concerns, appraisals, and contagiousness in a developmental context. *Developmental Review, 27*(1), 1-40. DOI: 10.1016/j.dr.2006.06.002

Steenbeek, H. W., & Van Geert, P. L. C. (2007b). 'Do you still like to play with him?' Variability and the dynamic nature of children's sociometric ratings. *Netherlands Journal of Psychology, 63*, 86 - 101.

Steenbeek, H., Van Geert, P., & Van Dijk, M. W. G. (2011). The dynamics of children's science and technology talents: A conceptual framework for early science education. *Netherlands Journal of Psychology, 66*, 96-109.

Steenbeek, H., Jansen, L., & van Geert, P. (2012). Scaffolding dynamics and the emergence of problematic learning trajectories. *Learning and Individual Differences, 22*(1), 64-75. DOI: 10.1016/j.lindif.2011.11.014

Steenbeek, H., & van Geert, P. (2013). The Emergence of Learning-Teaching Trajectories in Education: a Complex Dynamic Systems Approach. *Nonlinear dynamics psychology and life sciences, 17*(2), 233-267.

Steenbeek, H., van der Aalsvoort, D., & van Geert, P. (2014). Collaborative Play in Young Children as a Complex Dynamic System: Revealing Gender Related Differences. *Nonlinear dynamics, psychology, and life sciences, 18*(3), 251-276.

Sweller, J., Van Merriënboer, J.J.G., & Paas, F.G.W.C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296.

The Santa Fe Institute (2018). Glossary. Retrieved from https://www.complexity explorer.org/explore/glossary.

Trulla, L., Giuliani, A., Zbilut, J., & Webber, C. (1996). Recurrence quantification analysis of the logistic equation with transients. *Physics Letters A, 223*(4), 255-260. doi:10.1016/S0375-9601(96)00741-4

Van Dijk, M., & van Geert, P. (2007). Wobbles, humps and sudden jumps: A case study of continuity, discontinuity and variability in early language development. *Infant Behavior and Development, 16,* 7-33. DOI: 10.1002/icd.506

Van Dijk, M., & van Geert, P. (2015). The nature and meaning of intraindividual variability in development in the early life span. In M. Diehl, K. Hooker, & M. J. Sliwinski (Eds.), *Handbook of intraindividual variability across the life span* (pp. 37-58). Routledge, Taylor and Francis group.

Van Geert, P., & van Dijk, M. (2002). Focus on variability: New tools to study intra-individual variability in developmental data. *Infant Behavior & Development, 25,* 340-374.

Van Geert, P. (2003). Dynamic systems approaches and modeling of developmental processes. In J. Valsiner & K. J. Conolly. (Eds.), *Handbook of developmental psychology* (pp. 640-672). London, UK: Sage.

Van Orden, G.C., Holden, J.G., Turvey, M.T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology. General, 132*(3), 331-50.

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

Ward, L. M. (2002). *Dynamical cognitive science.* Cambridge, MA: The MIT Press.

Wijnants, M. L., Cox, R. F. A., Hasselman, F., Bosman, A. M. T., & Van Orden, G. (2012). A trade-off study revealing nested timescales of constraint. *Frontiers in Physiology, 3*.

Xu, Y., Yin, Y., & Yin, J. (2017). Tackling topic general words in topic modeling. *Engineering Applications of Artificial Intelligence, 62,* 124-133. doi:10.1016/j.engappai.2017.04.009

Appendix A

List of stopwords

"a"	"across"	"all"	"also"	"and"	"anyway"
"a's"	"actually"	"allow"	"although"	"another"	"anyways"
"able"	"after"	"allows"	"always"	"any"	"anywhere"
"about"	"afterwards"	"almost"	"am"	"anybody"	"apart"
"above"	"again"	"alone"	"among"	"anyhow"	"appear"
"according"	"against"	"along"	"amongst"	"anyone"	"appreciate"
"accordingly"	"ain't"	"already"	"an"	"anything"	"appropriate"
"are"	"associated"	"became"	"beforehand"	"best"	"by"
"aren't"	"at"	"because"	"behind"	"better"	"c"
"around"	"available"	"become"	"being"	"between"	"c'mon"
"as"	"away"	"becomes"	"believe"	"beyond"	"c's"
"aside"	"awfully"	"becoming"	"below"	"both"	"came"
"ask"	"b"	"been"	"beside"	"brief"	"can"
"asking"	"be"	"before"	"besides"	"but"	"can't"
"cannot"	"clearly"	"consider"	"couldn't"	"did"	"don't"
"cant"	"co"	"considering"	"course"	"didn't"	"done"
"cause"	"com"	"contain"	"currently"	"different"	"down"
"causes"	"come"	"containing"	"d"	"do"	"downwards"
"certain"	"comes"	"contains"	"definitely"	"does"	"during"
"certainly"	"concerning"	"corresponding"	"described"	"doesn't"	"e"
"changes"	"consequently"	"could"	"despite"	"doing"	"each"
"edu"	"entirely"	"everybody"	"except"	"followed"	"four"
"eg"	"especially"	"everyone"	"f"	"following"	"from"
"eight"	"et"	"everything"	"far"	"follows"	"further"
"either"	"etc"	"everywhere"	"few"	"for"	"furthermore"
"else"	"even"	"ex"	"fifth"	"former"	"g"
"elsewhere"	"ever"	"exactly"	"first"	"formerly"	"get"
"enough"	"every"	"example"	"five"	"forth"	"gets"
"getting"	"got"	"hardly"	"he's"	"hereafter"	"him"
"given"	"gotten"	"has"	"hello"	"hereby"	"himself"
"gives"	"greetings"	"hasn't"	"help"	"herein"	"his"
"go"	"h"	"have"	"hence"	"hereupon"	"hither"
"goes"	"had"	"haven't"	"her"	"hers"	"hopefully"
"going"	"hadn't"	"having"	"here"	"herself"	"how"
"gone"	"happens"	"he"	"here's"	"hi"	"howbeit"

"however"	"if"	"indicate"	"inward"	"its"	"kept"
"i"	"ignored"	"indicated"	"is"	"itself"	"know"
"i'd"	"immediate"	"indicates"	"isn't"	"j"	"knows"
"i'll"	"in"	"inner"	"it"	"just"	"known"
"i'm"	"inasmuch"	"insofar"	"it'd"	"k"	"l"
"i've"	"inc"	"instead"	"it'll"	"keep"	"last"
"ie"	"indeed"	"into"	"it's"	"keeps"	"lately"
"later"	"let's"	"looks"	"me"	"most"	"name"
"latter"	"like"	"ltd"	"mean"	"mostly"	"namely"
"latterly"	"liked"	"m"	"meanwhile"	"much"	"nd"
"least"	"likely"	"mainly"	"merely"	"must"	"near"
"less"	"little"	"many"	"might"	"my"	"nearly"
"lest"	"look"	"may"	"more"	"myself"	"necessary"
"let"	"looking"	"maybe"	"moreover"	"n"	"need"
"needs"	"no"	"not"	"of"	"on"	"other"
"neither"	"nobody"	"nothing"	"off"	"once"	"others"
"never"	"non"	"novel"	"often"	"one"	"otherwise"
"nevertheless"	"none"	"now"	"oh"	"ones"	"ought"
"new"	"noone"	"nowhere"	"ok"	"only"	"our"
"next"	"nor"	"o"	"okay"	"onto"	"ours"
"nine"	"normally"	"obviously"	"old"	"or"	"ourselves"
"out"	"particularly"	"presumably"	"r"	"regardless"	"same"
"outside"	"per"	"probably"	"rather"	"regards"	"saw"
"over"	"perhaps"	"provides"	"rd"	"relatively"	"say"
"overall"	"placed"	"q"	"re"	"respectively"	"saying"
"own"	"please"	"que"	"really"	"right"	"says"
"p"	"plus"	"quite"	"reasonably"	"s"	"second"
"particular"	"possible"	"qv"	"regarding"	"said"	"secondly"
"see"	"self"	"several"	"so"	"sometimes"	"specifying"
"seeing"	"selves"	"shall"	"some"	"somewhat"	"still"
"seem"	"sensible"	"she"	"somebody"	"somewhere"	"sub"
"seemed"	"sent"	"should"	"somehow"	"soon"	"such"
"seeming"	"serious"	"shouldn't"	"someone"	"sorry"	"sup"
"seems"	"seriously"	"since"	"something"	"specified"	"sure"
"seen"	"seven"	"six"	"sometime"	"specify"	"t"
"t's"	"thank"	"their"	"there's"	"these"	"third"
"take"	"thanks"	"theirs"	"thereafter"	"they"	"this"

"taken"	"thanx"	"them"	"thereby"	"they'd"	"thorough"
"tell"	"that"	"themselves"	"therefore"	"they'll"	"thoroughly"
"tends"	"that's"	"then"	"therein"	"they're"	"those"
"th"	"thats"	"thence"	"theres"	"they've"	"though"
"than"	"the"	"there"	"thereupon"	"think"	"three"
"through"	"took"	"trying"	"unless"	"use"	"v"
"throughout"	"toward"	"twice"	"unlikely"	"used"	"value"
"thru"	"towards"	"two"	"until"	"useful"	"various"
"thus"	"tried"	"u"	"unto"	"uses"	"very"
"to"	"tries"	"un"	"up"	"using"	"via"
"together"	"truly"	"under"	"upon"	"usually"	"viz"
"too"	"try"	"unfortunately"	"us"	"uucp"	"vs"
"w"	"we'd"	"were"	"whenever"	"whereupon"	"who's"
"want"	"we'll"	"weren't"	"where"	"wherever"	"whoever"
"wants"	"we're"	"what"	"where's"	"whether"	"whole"
"was"	"we've"	"what's"	"whereafter"	"which"	"whom"
"wasn't"	"welcome"	"whatever"	"whereas"	"while"	"whose"
"way"	"well"	"when"	"whereby"	"whither"	"why"
"we"	"went"	"whence"	"wherein"	"who"	"will"
"willing"	"would"	"you"	"yourself"		
"wish"	"would"	"you'd"	"yourselves"		
"with"	"wouldn't"	"you'll"	"z"		
"within"	"x"	"you're"	"zero"		
"without"	"y"	"you've"			
"won't"	"yes"	"your"			
"wonder"	"yet"	"yours"			

Appendix B

Domain-specific words

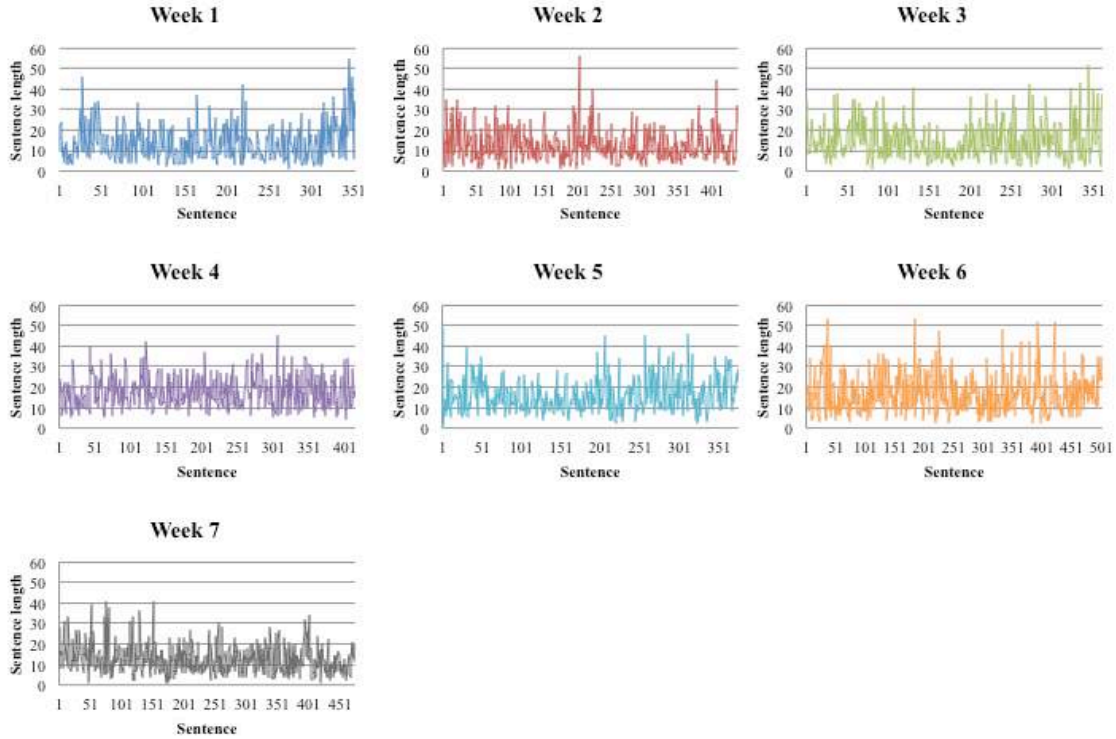
1/f noise	agent	Aristotle
adaptation	agent-based simulation	arms race
agent-based models (agent-based modelling)	alife (artificial life)	artificial intelligence
adaptive	allometry	asymmetry
adaptive system	ansatz	asynchrony
adaptive walk	anthropic principle	attractor
affine transformation	area-preserving map	autocatalysis
autonomous	bipartite graph	Church-Turing thesis
Baldwin effect	boundaries	chunking
autopoiesis (autopoeisis)	butterfly effect	clustering coefficient
basins of attraction	catastrophe theory	co-evolution
Benard system	cellular automata	coherence
betweenness centrality	canalization	community stability
bifurcation	cascading failure	community structure
bifurcation diagram	chaos	complex network
complex system	complex adaptive system	deterministic system
complexity	complex responsive processes	difference equation
conditonal entropy	cybernetics	differential equation
connectionism	cycle (graph theory term)	dimension (fractal dimension)
continuous time	Darwinism	directed graph
convergence	degree (of a node in a network)	discrete time
critical phenomena	degree distribution	double auction
criticality	deterministic dynamics	diversity
dissipative structure	emergent behavor (phenomena)	evolutionary developmental biology
dynamical system	enaction	evolutionary arms race
dynamic system	entropy	evolutionary game theory
economic equilibria	epigenesis	evolvability
ecophysics	equilibrium	exptation
effective complexity	ergodic process	fat tail
embodied cognition	ergodic theory	Feigenbaum's constants
emergence	ergodicity	far-from-equilibrium
feedback	game theory	heavy tail
fitness	genetic algorithm	heterogeneous agent models
fitness landscape	genetic drift	heterogeneity
food web	genotype	highly optimized tolerance
Fourier Transform	genotype-phenotype map	homeostasis
fractal	graph theory	highly optimized tolerance
Fractal landscape	Hamming distance	hypercycle

Game of Life	Hausdorff dimension	information
information theory	limit cycle	Lyapunov exponent
initial conditions	limit points	macroevolution
instability	link	macrostate
interaction	linear system	Markov chain
isomorphism	logistic map	maximum entropy
keystone species	long tail	metabolism
L-system	long-tailed distribution	microevolution
leverage point	low energy nuclear reactions	microstate
minimum specifications	nonlinear system	Pareto optimality
morphogenesis	nonlinearity	Parameters
mutual information	novelty	path dependence
network	NP	path length
network diameter	NP-complete	pathogen
neural networks	ontogeny	period (dynamical system term)
node	orbit	phase portrait
noise (signal)	P	phase space
phenotype	random walk	scale-free network
phylogeny	recurrence relation	second law of thermodynamics
positive deviance	redundancy	self-organization
power laws	repeller	self-organized criticality
principle of maximum entropy	resilience	self-similarity
public-goods games	robustness	sensitive dependence on initial conditions
punctuated equilibrium	saddle point	Shannon entropy
random network	scale	Shannon information
small world network (principle)	sub-linear growth	thermodynamics
social networks	super-linear growth	time series
stability	synchrony	tipping point
state (system theory term)	synthetic biology	top-down
state space	synchronization (sync)	tragedy of the commons
statistical mechanics	symbiogenesis	trajectory
stochastic process	system dynamics	Turing machine
strange attractor	swarmware and clockware	universal Turing machine
utility (economics)		
versatility		
von Neumann neighborhood		
white noise		
Zipf's law		

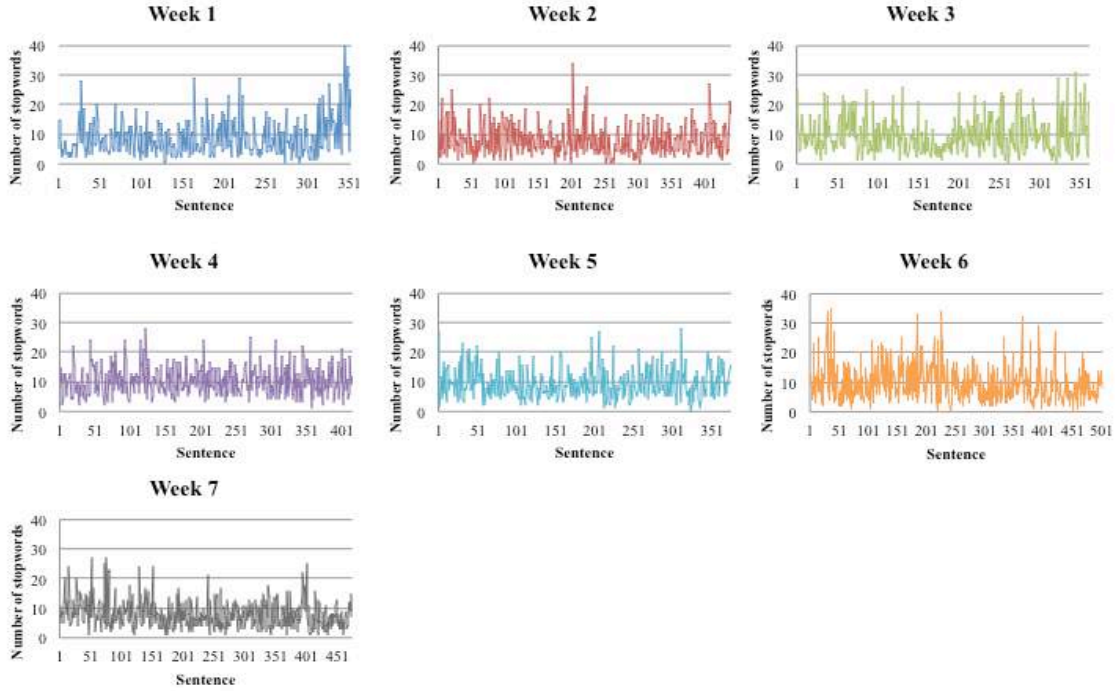
Appendix C

Time-series for each metric

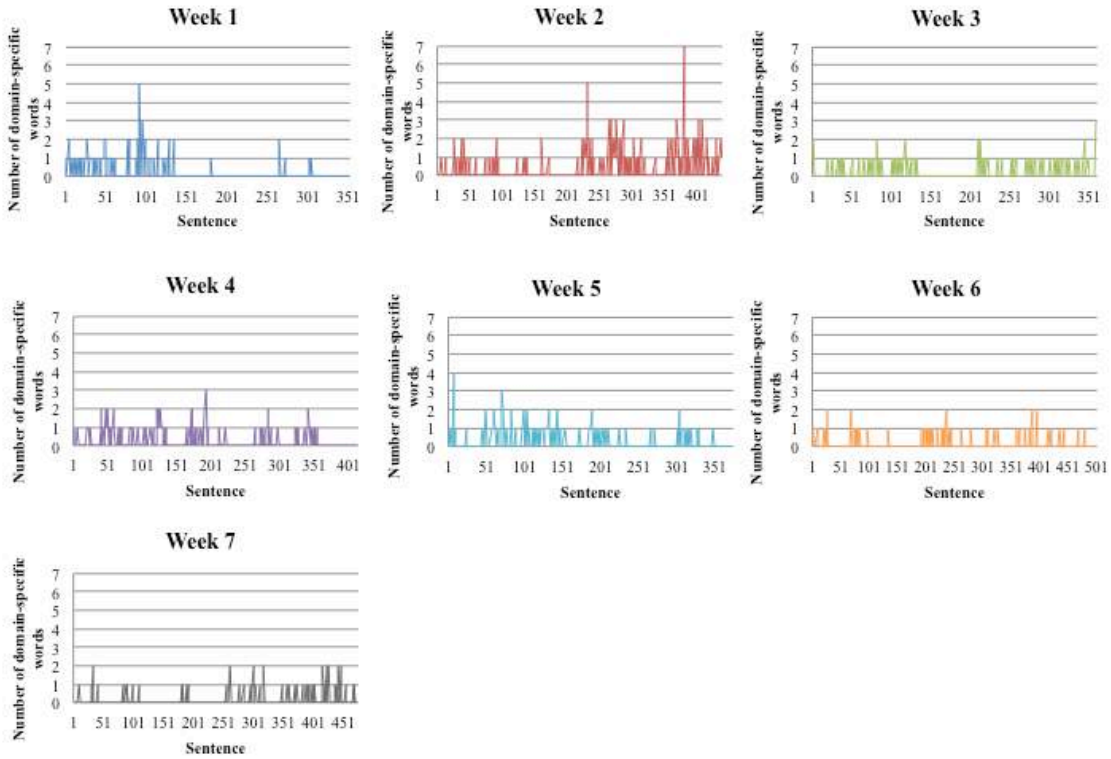
Sentence length



Stopwords



Domain-specific words



Appendix D

Average SL, SW, DSW, and Variability Scales for Each Metric in Each Lecture Video

Video	Average SL	Average SW	Average DSW	Variability scale SL	Variability scale SW	Variability scale DSW
1_6	12.97	7.55	0.71	NA	NA	NA
1_9	17.91	8.43	0.71	NA	NA	NA
1_13	13.48	8.59	0.50	-0.65	-0.34	-0.70
1_15	12.74	7.08	0.46	-1.21	-1.08	-0.68
1_18	11.78	7.90	0.02	-0.32	-0.44	-0.49
1_21	13.26	8.96	0.00	-0.68	-0.68	NA
1_25	11.21	6.94	0.07	-1.08	-0.78	-0.72
1_27	20.46	14.00	0.00	-0.51	-0.54	NA
2_2	15.17	9.33	0.25	NA	NA	NA
2_4	14.33	8.73	0.39	-0.50	-0.33	-0.55
2_6	11.33	6.56	0.11	NA	NA	NA
2_8	13.64	8.58	0.22	-0.90	-0.77	-0.37
2_10	12.72	8.32	0.07	-0.51	-0.53	-0.41
2_13	15.20	8.00	0.80	NA	NA	NA
2_14	10.67	6.13	0.69	-0.33	-0.44	-0.30
2_17	11.95	6.44	0.80	-0.26	-0.23	-0.18
2_19	10.27	7.48	0.27	-0.48	-0.51	-0.18
2_20	13.62	7.98	1.11	-0.89	-1.52	-0.29
2_23	13.48	9.48	0.66	NA	NA	NA
3_3	16.20	10.30	0.20	NA	NA	NA
3_4	13.25	8.54	0.13	NA	NA	NA
3_5	15.38	9.77	0.31	NA	NA	NA
3_6	17.27	10.36	0.18	NA	NA	NA
3_8	21.86	14.29	0.14	NA	NA	NA
3_9	14.56	8.63	0.30	NA	NA	NA
3_10	14.40	9.50	0.60	NA	NA	NA
3_11	16.20	10.87	0.60	NA	NA	NA
3_13	18.13	10.38	0.38	NA	NA	NA
3_14	10.29	6.68	0.00	-0.84	-0.67	NA
3_15	14.40	8.40	0.20	NA	NA	NA
3_17	15.47	10.23	0.20	-0.90	-1.07	-0.43
3_20	15.48	9.80	0.28	-0.39	-0.49	-0.54
3_21	15.00	10.02	0.26	-0.75	-0.55	-0.37
3_24	29.00	16.00	2.33	NA	NA	NA
4_2	15.78	9.42	0.24	-0.94	-1.11	-0.30
4_3	19.07	11.33	0.50	-0.76	-0.56	-0.41
4_6	17.27	9.91	0.33	-0.69	-0.78	-0.44
4_9	17.47	10.38	0.07	-0.46	-0.46	-0.34
4_13	18.83	10.51	0.30	-0.71	-0.80	-0.29
4_17	17.58	10.02	0.05	-0.86	-1.00	-0.59

5_2	16.61	10.55	0.23	-0.48	-0.56	-0.31
5_3	13.39	8.39	0.59	-0.58	-0.65	-0.41
5_7	12.87	8.23	0.23	NA	NA	NA
5_8	15.91	9.86	0.32	-0.32	-0.41	-0.15
5_11	12.80	8.13	0.07	NA	NA	NA
5_14	16.73	9.14	0.08	-0.37	-0.38	-0.47
5_16	16.96	9.39	0.18	-0.71	-1.02	-0.33
5_19	18.86	10.27	0.06	-0.28	-0.28	-0.46
6_2	18.40	11.64	0.13	-0.94	-0.59	-0.30
6_5	15.15	9.77	0.14	-0.20	-0.20	-0.19
6_9	18.58	11.85	0.07	-0.80	-0.59	-0.47
6_12	18.87	12.13	0.31	-1.23	-1.33	-0.63
6_14	15.40	8.63	0.21	-0.54	-0.57	-0.22
6_16	15.07	8.78	0.07	NA	NA	-0.64
6_18	14.50	8.41	0.07	-0.50	-0.61	NA
6_19	20.05	12.24	0.10	NA	NA	-0.67
6_22	16.86	8.84	0.24	-0.39	-0.48	NA
6_26	15.56	6.91	0.05	-0.32	-0.58	NA
6_28	20.64	9.91	0.09	NA	NA	NA
6_29	20.17	7.79	0.04	NA	NA	NA
6_31	28.25	11.00	0.00	NA	NA	NA
7_3	15.11	9.82	0.16	-0.54	-0.71	NA
7_6	14.02	9.18	0.06	-0.58	-0.48	NA
7_8	13.70	8.54	0.05	-1.09	-0.73	-0.68
7_11	10.73	6.73	0.03	-0.24	-0.16	-0.50
7_13	11.65	7.14	0.04	-0.36	-0.59	NA
7_15	12.27	7.40	0.25	-0.61	-0.71	-0.28
7_16	12.38	7.70	0.16	-0.59	-0.47	-0.55
7_17	12.78	8.13	0.25	-1.50	-0.64	-0.38
7_19	10.32	6.56	0.30	-0.48	-0.72	-0.35