LUNDS UNIVERSITET

INSTITUTIONEN FÖR PSYKOLOGI

EN: Cognitive Load Theory as a Predictor for Citation Count

SV: Cognitive Load Theory som en prediktor för antal citeringar

Samuel Ramos

Kandidatuppsats HT 2020

Handledare: Roger Johansson

Abstract

The purpose of this study is to examine if the cognitive load that is induced from different degrees of text complexity in scientific abstracts has an influence on how often the study will get cited. While the influence of cognitive load is well documented in research of activities such as attention allocation, decision making and reading comprehension, it is not known whether such effects can also be applied to more complex activities, such as whether a scientific paper will be influential (or not). In the present study, text induced cognitive load in scientific abstracts is captured by using the of Gunning Fog Index (GFI) and Dale-Chall Score (DCS) for text complexity, while also analyzing with the Type-Token Ratio (TTR). These three text analysis tools capture varying levels of text complexity from word length, sentence length, and lexical sophistication and diversity, which have been shown to impact cognitive load. Apart from examining this relationship per se, the present study also investigates if the three measures of text induced cognitive load varies between different academic fields. The study found TTR to be significantly correlated with the citation count in all three search results while DCS and GFI were not. A post hoc analysis revealed only the DCS as having a significant difference for each comparison within academic disciplines.

Keywords*:* Cognitive Load Theory, Gunning Fog Index, Type-Token Ratio, Dale-Chall Score

Tack!

Jag vill tacka min handledare Roger Johansson för alla goda råd, idéer och diskussioner som har lett fram till detta arbete och flera nya kunskaper. Jag vill också tacka Astrid Persson för hjälp med tillämpningen av de datavetenskapliga delarna i metodologin och figuren för flödesschemat.

Cognitive Load Theory as a Predictor for Citation Count

We are constantly exposed to large amounts of information in our everyday life by the technological demands of our society. This abundance of information is especially true within the scientific community, where the number of articles published is growing at nearly four percent yearly, with an estimated output of 2.6 million articles in 2018 (White, 2019). This increase in scientific literature makes processing information more difficult when researchers are limited by their investments of time and resources. Understanding how researchers are adapting to this shift in information availability can be important in predicting how the future of the scientific community will be influenced by this trend. There are many factors that influence the popularity of a research article such as the popularity of the journal, authors or subject matter. However, another factor that might affect the popularity of a scientific article might be in the structure of the text itself.

The purpose of this study is to examine if the cognitive load that is induced from different degrees of text complexity in scientific abstracts has an influence on how often the study will get cited. It is possible to evaluate the text complexity of many scientific papers by using software defined algorithms. In effect, such algorithms provide a numerical value of text complexity, which can be used as a proxy for the cognitive load a reader needs to invest when processing the text.

An example of how this can be viewed practically is how we as readers might choose to engage with online content or choose a book to read at a library. A younger audience will likely pick a book with many pictures and short words, while the older or more educated audience will gravitate towards reference books or difficult novels with a high page count. The selection criteria will feature sensory data (book thickness, pictures on the cover or word per page density) but it can also be dependent on how the language itself is used. It is possible that a book that fits what they expect of a book visually, but when they try to read it they find the words too long or the concepts too complex. This added layer of complexity will induce an increase in cognitive load and can create mental fatigue even though the book had a strong visual appeal. From this perspective, it could be predicted that the more cognitive load a scientific paper requires to be sufficiently read and understood, the lower the citation count.

The Processing Limitations of Human Cognition and Attention

The limitations of human cognition are an important part of cognitive psychology and theories of cognitive load that have been developed during the past century. For instance, when we are driving a car it is not possible for us to calculate all possible routes to a particular destination, even in familiar areas we might often struggle to relay all the details of each possible route. Especially if we are engaged in a conversation or listen to the radio at the same time. The human cognitive system's limitations are an important part of the research on working memory.

Working Memory

The term working memory is central to understanding cognitive load and can be defined as the set of mental processes which activates a limited amount of information for a temporarily accessible state. The notion that working memory is inherently limited was first suggested by G.A. Miller, whose experiments indicated that humans can generally hold seven plus or minus two units of information in such a temporary "storage" (Miller, 1956). This short-term memory (STM) was then developed as one of three stages by Atkinson and Shiffrin with their multi-store model of memory (Atkinson & Shiffrin, 1968). The STM became positioned between two separate memory stores: long term memory (LTM) and sensory memory (SM) (Atkinson & Shiffrin, 1968). The SM is composed of sensory registers for each sense, which are not consciously processed and merely hold the information for processing by the STM (Atkinson & Shiffrin, 1968). The sensory information that is selected by attention is then consciously processed in the STM (Atkinson & Shiffrin, 1968). In the third stage the information processed by the STM is then encoded into the LTM, where it is generally held to have unlimited storage (Atkinson & Shiffrin, 1968).

The investigations of the link between STM and LTM would lead to the development of a new "multicomponent model" proposed by Alan Baddeley and Graham Hitch (Baddeley, 2012). Although STM and WM are sometimes used interchangeably, the Baddeley model would use STM to refer to storage, while WM combined both storage and manipulation (Baddeley, 2012). This model defined a WM to comprise of three components: the visuo-spatial sketchpad, the episodic buffer, and the phonological loop (VSS, EB, and PL respectively) (Baddeley 2012). The VSS refers to the visual information consciously being processed, the PL would process auditory information and the EB would mediate between these two conscious processes (Baddeley, 2012). These three processes of working memory are actively moderated by a central executive (CE) which can shift focus between the processing of, for instance, auditory and visual information (Baddeley, 2012). The Baddeley model would also further distinguish between a fluid system (CE, VSS, EB, and PL) which temporarily activates these functions and a crystallized system which has permanent skills and knowledge (Baddeley, 2012). This crystallized system (analogous to LTM) is comprised of areas corresponding to the fluid system: visual semantics, episodic LTM, and language (Baddeley, 2012). There are several characteristics of the phonological loop which have shown limiting effects on the working memory. First that there is a phonological similarity effect, where similar sounds are harder to memorize than dissimilar sounds (Baddeley, 2012). Secondly, a word length effect, where it is shown that one syllable word sequences are easier to commit to process and maintain than sequences which contain polysyllabic word sequences (Baddeley, 2012). Thirdly, there is an articulatory suppression effect where continuous subvocalization of a single word can affect the processing and recall of information (Baddeley, 2012). The predicament of such language-related WM effects is an important part of the analysis of this study.

Recent investigations have given rise to other conceptualizations of WM. Research by Nelson Cowan has proposed that instead of the five plus or minus two limit by Miller, it is more likely that the WM capacity is limited to three to five meaningful items (Cowan et al., 2010). Cowan also advocates for a limited capacity WM that is based on the idea that we can temporarily activate a limited amount of specific information from LM, rather than having specialized buffers for certain types of information (as in the Baddeley model) (Cowan, 2005). Nevertheless, common to all current models of WM is that there is a limitation of how much information we can dedicate our attentional resources towards, which is integral for the application of cognitive load theory (CLT).

Working Memory and CLT

The two important limitations of working memory that influence CLT are that it is limited in capacity (Miller, 1956) (Cowan et al., 2010) and in duration (Peterson & Peterson, 1959). These limitations of working memory capacity are important to CLT as they concern factors which can slow down or speed up a learning process. CLT can be divided into three separate parts: intrinsic, extraneous, and germane loads (Kirschner, 2002). The intrinsic cognitive load describes the inherent difficulty that is associated with a particular task (Sweller, 1994). An example of this could be seen by looking at a simple chemical reaction such as combining sodium chloride and water, in comparison to the protein folding that occurs in the intracellular matrix of a human cell. The first example can be explained easily because it involves very elementary processes, whereas protein folding can be so difficult that it takes researchers and machine learning years to solve. The extraneous factor accounts for information which is influenced by external factors outside of the demands of the task itself (Sweller, 1994). Another way to understand extraneous factors is that these are irrelevant information sources which have a *negative effect* on the learning process. This could be described as when a student is learning information on algebra and at the same time is having a friend trying to discuss while the professor is talking. A germane load describes the process by which an individual creates schemas to process information more efficiently (Van Merriënboer & Sweller, 2010). These factors *positively influence* the learning process and aid in the storing of information into long term memory.

The inherent difficulty of a particular task within CLT is determined by the amount of element interactivity (Debue & Leemput, 2014). An example of this would be learning the locations of countries on a world map, as each country is matched to a name and a shape/location. Whereas learning relationships described by university level mathematics will require many pre-learned concepts to be applied before attempting to solve any of the equations. The discipline of reading can involve a hierarchy of elements which take years of experience to acquire the necessary knowledge to integrate the concepts presented in the reading material. This developmental and hierarchical nature of reading is useful for evaluating the cognitive load of a written text.

Text Comprehension and Cognitive Load

Text comprehension is processed within the limited working memory framework. A model of cognitive load generated by media specific reading comprehension has been proposed by Zumbach and Mohraz which analyses text from three perspectives: complexity of content, narration format and text design (Zumbach & Mohraz, 2008). The model assumes that text or hypertext learning environments vary in cognitive load based on problems surrounding orientation. Therefore, disorientation can be a consequence of; unstructured texts (e.g., missing text), the content itself being overly complex or environmental conditions that affect the reader (Zumbach & Mohraz, 2008). Disorientation can also be the result of limitations surrounding the capacity of the individual (Zumbach & Mohraz, 2008). The authors also note that media literacy and prior knowledge of the individual can affect the cognitive load. Further, they attributed content complexity as a condition which increases intrinsic load, while narration and text design act as extraneous loads. The particular focus of this study is on the complexity of the content and the readability of the text itself. The algorithms used to capture such aspects are further described in the methods section on their contributions to text induced cognitive load.

Previous Research on Text Induced Cognitive Load

To my knowledge, no previous study has examined the popularity of scientific literature from a cognitive load perspective. However, there are several studies that examine the links between cognitive load and reading tasks. Reading task complexity can have an impact on the cognitive load and affect the learning process (Ghanimi et al., 2016). It has been shown that there is an effect of increased cognitive load on individuals task performance when evaluating the influence of hypertext on readers (DeStefano & Lefevre, 2007). The article by Stefano reviews the literature surrounding the effect of hypertext on reading and concludes that hypertext is shown to increase the cognitive load. This was measured by looking at 38 studies over a 14-year period which covered hypertext navigation and comprehension regarding how these features influence cognitive demands. Sentence length (used in two of the algorithms in this research) has also shown impact on cognitive load (Mikk, 2008). The study examined an optimal sentence length for 17- to 18-year-old students, by using 30 cloze tests (Mikk, 2008). The study found that instead of a linear relationship between sentence length and cognitive load there was an optimal zone between 130-50 characters (Mikk, 2008).

While there are previous studies which have researched the link between cognitive load and text processing, there is currently no literature that has looked at the implication of this effect on the scientific popularity of a research article.

Aims and Objectives

- The purpose of this study is to examine if text induced cognitive load from scientific abstracts has an influence on how often a study gets cited. It is predicted that higher text induced cognitive load will correlate negatively with how often a study has been cited.
- An additional aim is to investigate if the text induced cognitive load from scientific abstracts varies between academic disciplines.

Method

Design

This research uses a correlational study design where publicly available journal article abstracts are evaluated by three linguistic algorithms. These algorithms create quantitative evaluations of the abstracts which are then measured against the citation count to see if a linear relationship exists. The purpose of using three algorithms is to provide multiple analysis of text complexity which affect the difficulty of reading comprehension and therein the cognitive load. The study design also includes three separate search criteria that were used to request data from the Microsoft Academic database: “Personality Trait Theory”, “Social Contract Theory”, and “Coagulase Test”. This research was limited to only sample a subset of the research papers available from the Microsoft Academic Database, due to processing power required to analyze the entire database. These three search criteria analyze subject matters from different disciplines within the scientific field to test if the studied relationship exists in all three search results. This is due to the possibility that specific areas of study might exhibit different typical writing styles. Other research by Plavén-Sigray using the Flesch Reading Ease (FRE) and New Dale-Chall formulas has shown differences between different scientific disciplines in the mean slope as determined by the text complexity (Plavén-Sigray et al., 2017). This study will test if these differences occur between the three scientific disciplines utilized. It is not predicted that an effect of cognitive load generated by text complexity would be limited to a single area of research, therefore three separate areas have been selected for study to either enhance or limit the results of one search query. Each search field represents a topic within a different academic discipline which have been chosen at random, see Table 1, that has more than 100 and less than 1000 results. This study uses public domain algorithms which are freely available online so that other researchers may use this library to apply the analysis. The code used in this research will be made available on GitHub so that the methods can be reapplied, provided the researcher is using the same data inputs.

Table 1. Scientific Disciplines for the Searches

Search request	Scientific discipline
“Personality Trait Theory”	Psychology

"Social Contract Theory"	Sociology
"Coagulase Test"	Medicine

Assessments and Measures

There are three algorithms that have been chosen for this particular study: Gunning Fog Index (GFI), Dale-Chall Score (DCS), and Type-Token Ratio (TTR).

Gunning Fog Index. The GFI is calculated by the formula below:

$$0.4\left[\left(\frac{\text{words}}{\text{sentences}}\right)+100\left(\frac{\text{complex words}}{\text{words}}\right)\right]$$

This algorithm conceptualizes two aspects of the English language which affects the complexity of a written research article: the number of words that are in each sentence and the number of "complex words" that are contained within the writing passage. Another simpler version of writing this formula is: "G = 0.4(S + W)" with S representing sentence length and W representing the percentage of words with three or more syllables. This algorithm is rather popular in determining the reading difficulty of a particular passage.

Dale-Chall Score. There are many words in the English language that possess quite a few syllables and yet are not considered difficult or are widely well known. Conversely, there can be words with three or less syllables that are considered esoteric in the modern vernacular. With this constraint in mind, it was important to choose another algorithm that could also examine text complexity but with a stronger emphasis on verbal complexity.

$$0.1579\left(\frac{\text{difficult words}}{\text{words}}\times 100\right)+0.0496\left(\frac{\text{words}}{\text{sentences}}\right)$$

While the sentence length component is fairly similar to the GFI, the word complexity measurement is different. This algorithm uses a list of 3000 familiar words (Chall, 1995) which are weighted against the total word count to create a value which shows the relative difficulty of a particular reading passage. The purpose of the DCS is to compensate for the simple measure of word complexity in the first algorithm, while offering a secondary measurement of the same texts.

Type-Token Ratio. The final algorithm that is used is the TTR, which is used as a measurement of lexical diversity. Lexical diversity offers a secondary perspective on the complexity of a particular text in that it looks at the total number of unique words (types) divided by the total number of words (tokens) (Johansson, 2008). This algorithm (unlike the other two used in this research) dictates that the abstracts be tokenized before being processed, which transforms the abstracts into lists of the words they contain. The process of transforming the abstract's text into a list of words allows the algorithm to not be subjected to errors created by bad formatting. The higher the value is equivalent to a higher level of vocabulary diversity with a lower level indicating that there exists a substantial amount of repetition within the text.

Algorithm Summary

The GFI and DCS algorithms both analyze text complexity from a readability perspective. These two algorithms give an indication of how difficult the words and sentences are to parse through. They do not however account for other measures of complexity and can also assume that because a word is long it would be inherently difficult, or if it is short it is inherently less difficult. The GFI algorithm does not factor this at all, while the DCS does factor in a 3000-word subtraction process based on commonality of the words. The TTR accounts for complexity differently than the first two algorithms by measuring the duplication of words against the total text. The first two algorithms do not account for the words in relationship to each other only their character length and sentence length. Table 2 gives an overview of the range of values that might be achieved with the different algorithms and how to interpret them.

Table 2

Values for the Readability Metrics

Gunning Fog Index		Dale-Chall Score		Type-Token Ratio	
17	Post-University	10+	Post-University	1.0	High lexical diversity
16	University Senior	9	Freshman-Junior		
15	Junior	8	11th to 12th Grade		
14	Sophomore	7	9th to 10th Grade		
13	Freshman	6	7th to 8th Grade		

12	12th Grade	5	5th to 6th Grade		
11	11th Grade	4	4th Grade and below		
10	10th Grade				
9	9th Grade			0.01	Low lexical diversity

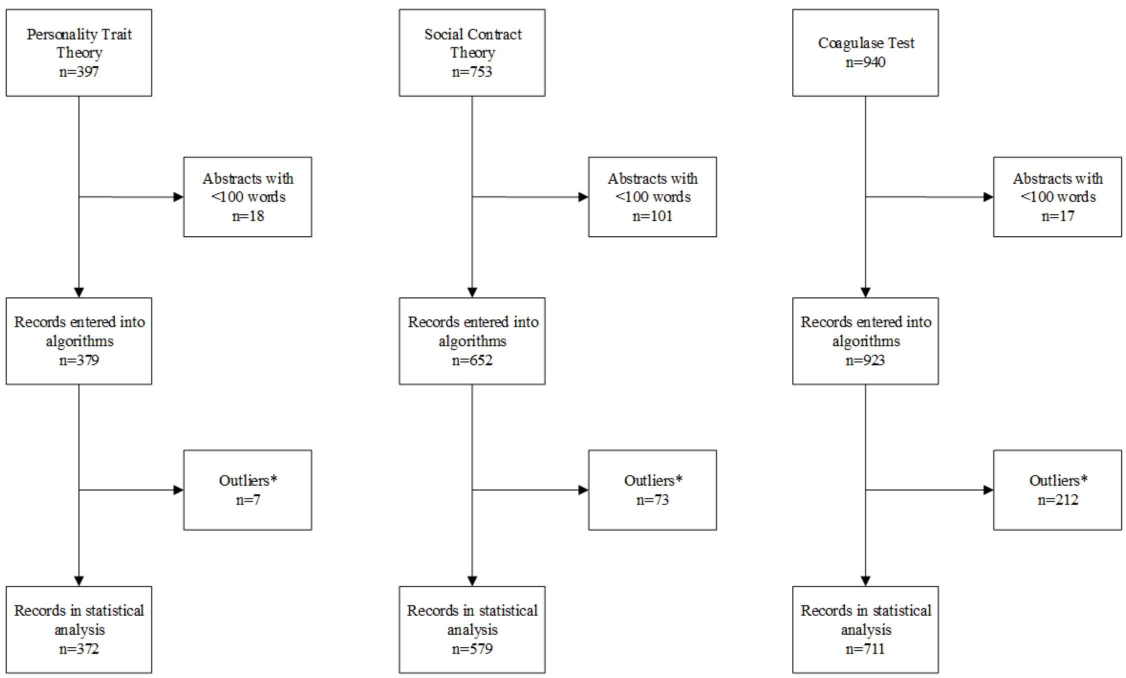
Note. The left-sided columns contain the values of the respective readability metric and the right-sided columns the matching approximate education level.

Procedure

The “Publish or Perish” software version 7.27.2949 (Harzing, 2007) was used to download research publication data from the Microsoft Academic database (Sinah, 2015). Several criteria were used in filtering the abstracts into a dataset that was usable for the purpose of this study. First, the search criteria were used to collect a random assortment of research articles surrounding a particular topic of interest (“Personality Trait Theory”, “Social Contract Theory” and “Coagulase Test”). Secondly, these search terms were filtered for results only within the range of years between 2010 and 2015. Thirdly, the search results were then filtered by type which excluded all types outside of journal articles (i.e. patents, books, conference papers etc.). The final step was a manual reading of the article abstracts to find and exclude non-English articles from the study, as these would not be usable for the word complexity tests that are being applied. After these filters were applied the data was then exported as a comma-separated values file into the integrated development environment PyCharm Community Edition version 2020.3 (JetBrains, 2020), where the final data sorting procedure and application of the algorithms was done using the high-level programming language Python version 3.6 (Van Rossum, 2009). For calculating GFI and DCS, the py-readability-metrics package version 1.3.5 (DiMascio, 2020) was used and for TTR, the lexical-diversity package version 0.1.1 (Kyle, 2020) was used.

As a first task, the abstracts were analyzed so that they were an acceptable word count for both the GFI and the DCS algorithms to be performed, which required a minimum of 100 words. Within the three search criteria: “Personality Trait Theory”, “Social Contract Theory”, and “Coagulase Test” there were 18, 101, and 17 abstracts which were affected by this. The abstracts were then analyzed by the GFI, DCS, and TTR. The TTR requires an extra data processing step over GFI and DCS, in that the algorithm requires the data be tokenized in order to be processed. A second filter was applied to abstracts which had GFI and DCS values that were influenced by formatting issues which produced values which are far outside the normal ranges produced by properly formatted abstracts. An upper limit of a GFI score of 40 and DCS score of 20 was set as a criterion to filter these results from the study. Within the three search criteria "Personality Trait Theory" (PTT), "Social Contract Theory" (SCT), and "Coagulase Test" (CT) there were 7, 73, and 212 abstracts which were affected by this (Figure 1).

Figure 1. Abstract Flow Diagram



* Gunning Fog Index >40 and/or David-Chall Score >20

The statistical analyses were conducted in SPSS version 25.0. Microsoft Visio Professional 2016 version 1901 was used to create the flowchart and the Python package matplotlib version 3.3.3 (Hunter, 2007) was used for creating the plots in PyCharm.

Results

Multiple regression was used to test if the measures of GFI, DCS, and TTR were predictors of citation count. The citation data for the three search results showed a highly skewed distribution which required the application of a logarithmic transformation (see appendix A). The logarithmic transformation of the citation count will henceforth be the reference for all “citation count” data. The presence of multicollinearity and residual normality was not detected to be of significance in deviation from the assumptions of the model (see appendix B).

Descriptive Statistics

The “Personality Trait Theory” search criterion had a mean GFI of 20.4 with a standard deviation of 3.3 (M = 20.4, SD = 3.3). The mean DCS was 12.3 with a standard deviation of 1.1 (M = 12.3, SD = 1.1). The mean TTR was 0.5 with a standard deviation of 0.1 (M = 0.5, SD = 0.1). The central tendency used for the median citation count was 8.0 (*Median* = 8.0) with 20.4% of the papers having no citations.

The “Social Contract Theory” search criterion had a mean GFI of 20.4 with a standard deviation of 3.3 (M = 20.4, SD = 3.3). The mean DCS was 11.8 with a standard deviation of 1.1 (M = 11.8, SD = 1.1). The mean TTR was 0.5 with a standard deviation of 0.1 (M = 0.5, SD = 0.1). The central tendency used for the median citation count was 1.0 (*Median* = 1.0) with 42.3% of the papers having no citations.

The “Coagulase Test” search criterion had a mean GFI of 19.0 with a standard deviation of 4.2 (M = 19.0, SD = 3.3). The mean DCS was 13.4 with a standard deviation of 1.3 (M = 13.4, SD = 1.3). The mean TTR was 0.5 with a standard deviation of 0.1 (M = 0.5, SD = 0.1). The central tendency used for the median citation count was 2.0 (*Median* = 2.0) with 36.1% of the papers having no citations (See Appendix C for all descriptive statistics, including minimum and maximum values.).

Results from Multiple Linear Regression

Assessing Assumptions

The assumptions of normality and residual normality were violated for the citation count data in all three search criteria, which was no longer seen after the data was log (base 10) transformed (Figure 2-7). There were no violations regarding the assumption of multicollinearity, where all variance inflation factors were below recommended level of five (Table 4).

Table 4

Variance Inflation Factors for Each Search Criteria and Readability Metric

	Gunning Fog Index	Dale-Chall Score	Type-Token Ratio
Personality Trait Theory	1.19	1.20	1.01
Social Contract Theory	1.24	1.30	1.06
Coagulase Test	1.15	1.17	1.02

“Personality Trait Theory” Results

Multiple regression analysis indicated that the three predictors significantly explained 5.9% of the variance (R^2 = .059, F(3,292) = 6.12, p < .001). The DCS and TTR variables significantly predicted the number of citations with a negative correlation (β = -.157, p = .012) and a positive correlation (β = .152, p = .008), respectively, and with GFI (β = -.044, p = .474) being nonsignificant (Table 5).

“Social Contract Theory” Results

Multiple regression analysis indicated that the three predictors significantly explained 4% of the variance (R^2 = .040, F(3,330) = 4.64, p = .003). TTR significantly predicted the number of citations with a positive correlation (β = .147, p = .009), however GFI (β = -.034, p = .569) and DCS (β = .118, p = .055) showed no significance (Table 5).

“Coagulase Test” Results

Multiple regression analysis indicated that the three predictors explained 1.1% of the variance, without significance (R^2 = .011, F(3,450) = 1.68, p = .17). Only TTR significantly predicted the number of citations with a negative correlation (β = -.096, p = .045) with both GFI (β = -.048, p = .341) and DCS (β = .021, p = .673) being nonsignificant (Table 5).

Table 5

Multiple Linear Regression Results

	R^2	p	β	p
Personality Trait Theory	.059*	.001		
Gunning Fog Index			-.044	.474
Dale-Chall Score			-.157*	.012

Type-Token Ratio			.152*	.008
Social Contract Theory	.040*	.003		
Gunning Fog Index			-.034	.569
Dale-Chall Score			.118	.055
Type-Token Ratio			.147*	.009
Coagulase Test	.011	.170		
Gunning Fog Index			-.048	.341
Dale-Chall Score			.021	.673
Type-Token Ratio			-.096*	.045

*Significant at the 0.05 level (2-tailed)

Differences in Readability Between Academic Disciplines

Gunning Fog Index

A one-way between subjects ANOVA was conducted to compare the effect of GFI on citation count for the three search topics [$F(2, 1659) = 23.43$, $p < .001$]. Post hoc comparisons using the Tukey HSD test indicated that the mean score for the CT condition ($M = 19.04$, $SD = 4.19$) was significantly different from the PTT condition ($M = 20.36$, $SD = 3.32$). There was also a significant difference between the CT and the SCT ($M = 20.41$, $SD = 4.11$) conditions. However, there was no significant difference between the PTT and SCT condition.

Dale-Chall Score

A one-way between subjects ANOVA was conducted to compare the effect of DC on citation count for the three search topics [$F(2, 1659) = 302.95$, $p < .001$]. Post hoc comparisons using the Tukey HSD test indicated that the mean score for the CT condition ($M = 13.41$, $SD = 1.34$) was significantly different from the PTT condition ($M = 12.28$, $SD = 1.08$). There was also a significant difference between the CT and the SCT ($M = 11.77$, $SD = 1.14$) conditions as well as the PTT and SCT conditions.

Type-Token Ratio

A one-way between subjects ANOVA was conducted to compare the effect of TTR on citation count for the three search topics [$F(2, 1659) = .021$, $p = .029$]. Post hoc comparisons using the Tukey HSD test indicated that the mean score for the CT condition ($M = 0.54$, $SD = .068$) was not significantly different from the PTT condition ($M = 0.54$, $SD = .086$). However, there was also a significant difference between the CT and the SCT ($M = 0.55$, $SD = .080$) conditions. There was no significant difference between the PTT and SCT condition.

Correlation Between Readability Metrics Within Each Search Criterion

There were significant correlations between GFI and DCS and DCS and TTR within all search criteria. The Pearson's correlation coefficients were for "Personality Trait Theory" $r(370) = .40$, $p < .001$ for GFI and DCS and $r(370) = .14$, $p = .005$ for DCS and TTR (Table 6); "Social Contract Theory" $r(577) = .48$, $p < .001$ for GFI and DCS and $r(577) = .16$, $p < .001$ for DCS and TTR (Table 7); and "Coagulase Test" $r(709) = .49$, $p < .001$ for GFI and DCS and $r(709) = .18$, $p < .001$ for DCS and TTR (Table 8). Thus, there was a moderate degree of positive correlation between GFI and DSC and a low degree of positive correlation between DCS and TTR within all search criteria. A significant correlation between GFI and TTR was only seen within "Coagulase Test", which was a low degree of positive correlation with a correlation coefficient of $r(709) = .13$, $p < .001$.

Table 6

Pearson's Correlation Coefficients Between Readability Metrics for Personality Trait Theory

	Gunning Fog Index	Dale-Chall Score	Type-Token Ratio
Gunning Fog Index	1.00	.40*	.09
Dale-Chall Score	.40*	1.00	.14*
Type-Token Ratio	.09	.14*	1.00

*Significant at the 0.05 level (2-tailed)

Table 7

Pearson's Correlation Coefficients Between Readability Metrics for Social Contract Theory

	Gunning Fog Index	Dale-Chall Score	Type-Token Ratio
Gunning Fog Index	1.00	.48*	.03
Dale-Chall Score	.48*	1.00	.16*
Type-Token Ratio	.03	.16*	1.00

*Significant at the 0.05 level (2-tailed)

Table 8

Pearson's Correlation Coefficients Between Readability Metrics for Coagulase Test

	Gunning Fog Index	Dale-Chall Score	Type-Token Ratio
Gunning Fog Index	1.00	.49*	.13
Dale-Chall Score	.49*	1.00	.18*
Type-Token Ratio	.13*	.18*	1.00

*Significant at the 0.05 level (2-tailed)

Discussion

The Relationship Between Cognitive Load and Citation Count

The result of the multiple linear regression failed to predict a change in citation count based on the combined values of the GFI, DCS and TTR. The results of the study showed TTR to be a predictor of citation count ("Personality Trait Theory": $\beta = .152$, "Social Contract Theory" $\beta = .147$ and "Coagulase Test": $\beta = -.096$). While the reported effect might seem small, it is unlikely that a large effect would occur as a result of this cognitive load effect. The study did not control for other important factors such as general impact of the journals or the authors involved. These factors are more likely to play a larger role in citation count than the text induced cognitive load factors studied here.

TTR also only showed a negative effect in relation to the intrinsic cognitive load *in one of the three* cases. The research hypothesis would have predicted in each case that a negative correlation existed for the increased cognitive load effect. The study found no significant linear correlation between the citation count and the GFI or DCS algorithms. A possible explanation for the difference between the GFI/DCS and the TTR results might be in how they had different approaches to data analysis. The TTR requires that the abstracts be "tokenized" and is thus not affected by issues with format errors (see the method section for more details on this process). It is also possible that the GFI and DCS metrics did not show the same effect because they do not account for lexical diversity in contrast to TTR. While the TTR showed a significant correlation in the three searches above, the GFI and DCS metrics likely need to be examined further with better controls to further test their influence as a predictor for citation count.

While the one-way ANOVA revealed that each text-induced cognitive load metric had significant differences between academic disciplines, the post hoc analysis only showed the DCS as having a significant difference for each comparison within academic disciplines. Only the CT/SCT pair showed a significant difference across all text-induced cognitive load metrics. However, the post hoc analysis did not reveal that any of the search criteria had a consistently higher value across all three text-induced cognitive load metrics. The correlational testing revealed that the text-induced cognitive load metrics were reasonably consistent in that they showed significant positive correlations within each search criteria, in all but two out of the nine cases. While the Plavén-Sigray study has shown different mean slopes over time between different scientific disciplines, the results of this study indicates that for the three different scientific disciplines there was a significantly different average mean value for text complexity during the five-year time frame (Plavén-Sigray et al., 2017). This is in line with the previous research which has indicated differences in mean text complexity using the Flesch Reading Ease and new DCS algorithms between different scientific disciplines (Plavén-Sigray et al., 2017).

Scientific Literature and Text Complexity

The relationship between the text analysis and social behavior is a growing field of cognitive science, psycholinguistics and data science. A common perspective on a research article is often based on the context or h-index, but modern readability analysis has made it much easier to examine the writing structure of multiple articles. An analysis of these factors allows for a more subtle examination of what affects the popularity of a journal article. The DCS, GFI or TTR algorithms can be applied to a diverse range of written texts or spoken passages. When applying this analysis to the complexity of presidential speeches it's been found that the text complexity has decreased from an university level of education (early 1800's) to a modern standard of around an eighth grade reading level (Thompson, 2014). Similarly, it has been found that an analysis of congressional speeches also shows a downward trend, with an average speaking level of 11.5 in 2005 lowering to 10.6 in 2012 (Drutman, 2012). It's not necessarily the case that text complexity is becoming simpler, as a study of 709,577 scientific abstracts has shown it is harder to read scientific papers now than in the past (Plavén-Sigray et al., 2017). The trends depicted by these longitudinal studies illustrate a shift in language usage and could be complemented by scientific inquiry as to which psychological processes might be shaping these trends. The readability of this article's abstract was a GFI of 19.9 and DCS of 9.4.

Cognitive Load Implications

The design of this study has accounted for specific text properties which have shown contributions to cognitive load (word length, sentence length and lexical diversity). This only measures those text-based factors as a measurement of intrinsic cognitive load, which may be less relevant than other cognitive load factors introduced by text complexity. Other factors surrounding text complexity can include visual distractions to the text presentation (extraneous cognitive load) or visual aids which can act as schema (germane) for the aid of processing the text. Other studies have shown that hypertext and even font size can be affected by or affect the cognitive load on processing text (Zumbach & Mohraz, 2008) (Luna et al., 2019).

The results of this study add to the current literature surrounding text complexity, cognitive load and online media. A study by Luna, Albuquerque, and Martin-Luengo examined the cognitive load effect (determined by sentence length) and showed that it altered the expected behavioral response of the participants judgement of learning (JOL). While the participants reported an expectedly higher JOL for text written in a larger font size, this effect diminished in relation to the increased cognitive load (Luna et al., 2019). My current study design has not accounted for the differences in font size; however, it is possible that it could be another factor which also influences citation count.

Another study which also associates cognitive load with text complexity is addressed by Mohraz and Zumbach (Zumbach & Mohraz, 2008). This study looks at the effect of hypertexts (as linear and nonlinear narratives), which are factored as an extraneous cognitive load on the readability of various texts. They also factor in the complexity of the text itself as a contrast to the role of hypertext as an extraneous cognitive load. My study does not account for the role of hypertext induced cognitive load in relation to citation count. Another component of intrinsic load that may also be useful to control for, is the level of education or familiarity with a particular topic of an individual as it may influence citation count.

An analysis of hypermedia using germane load measured against intrinsic and extraneous CL has been utilized in a study by Debue (Debue & Leemput, 2014). The results of which showed an opposite expected relationship between germane and extraneous and a nonlinear association between intrinsic and germane (Debue & Leemput, 2014). These findings illustrate that the role of a germane load in text comprehension can be difficult to study methodologically. However, the purpose of my study is only to examine the role of intrinsic cognitive load constrained to factors surrounding text complexity. It could be beneficial to use a controlled study approach to examine the text complexity factors in relation to the extraneous and germane loads that can also affect the popularity of a scientific article. This type of study design could pair well with the text complexity analysis provided in this study. There is a potential that graphical displays and variations on the different sites where the articles are hosted, have an increased extraneous load effect on the processing of text by working memory.

The effects of text comprehension on CL are only a single factor in CLT while extraneous and germane factors also control the effect on cognitive load. The studies above illustrate how extraneous CL (through hypertexts as nonlinear narratives) and germane CL (self-reported using a 10-point Likert scale) can also have an effect on the comprehension of text-based media (Zumbach & Mohraz, 2008) (Debue & Leemput, 2014).

Potential Design Improvements for Future Research

There are several design improvements that can be made to the study design to more accurately test for the possible effect of cognitive load on the citation count.

Using a Controlled Study Design. A controlled study design could test this hypothesis with greater accuracy. If the abstracts were written at various levels of difficulty and graduate students were to choose which abstracts, they found more useful relative to others, it could show in a controlled setting whether the level of text complexity affected their likelihood to cite a particular article.

Controlling for Confounding Variables. An issue with the current study design is that it does not account for the impact of the journals or the authors. These factors are likely to play a role in the citation count, which will affect the study of cognitive load based upon text complexity. A possible solution for this could be using only a select journal abstracts over a longer period and adjusting for time since original publishing. This also helps to control for formatting issues that exist in smaller journals which are hosted on the search engines data aggregation. The impact of authorship and journal has been studied by using author impact factors (AIF) and journal impact factors (JIF) and could be applied to the data that was collected in this study (Amjad et al., 2019) (Garfield, 2006).

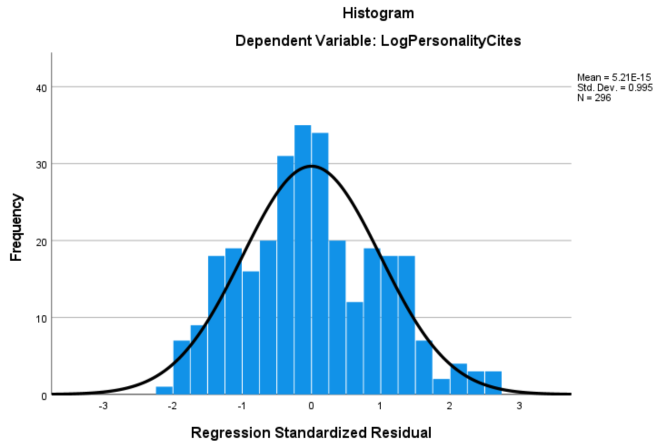
Using a Larger Sample Size. The search criteria listed above is a relatively small sample size compared to the total number of published research articles that is available on those databases.

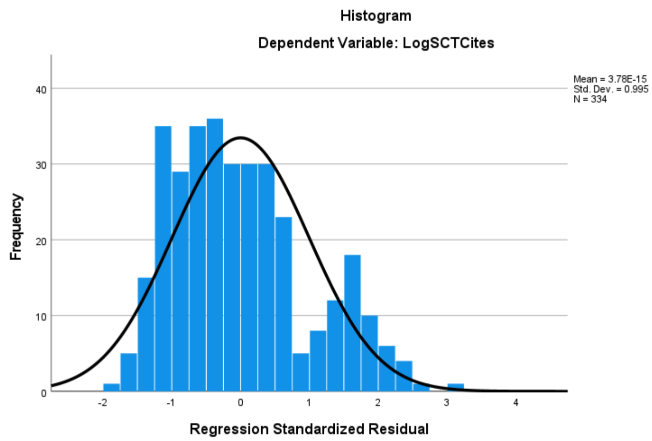
The Cognitive Load Effect May Not Be Linear. It is very possible that when accounting for the above that the hypothesis of the cognitive load prediction is overly simplistic. The cognitive load may be negatively correlated only after a certain threshold, while the research articles which are written at a lower-than-average level of text complexity show a positive correlation. A similar effect was found in the Mikk study where the author found that instead of a linear relationship between sentence length and cognitive load there was an optimal zone between 130-50 characters (Mikk, 2008).

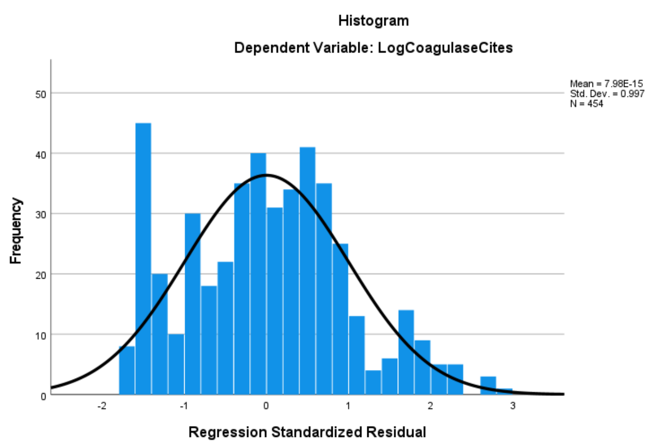
Expanding the Analysis. It might also help to expand the data being studied, from examining only the abstract and instead looking at the entire research article. It is important to note that the method employed by this study would likely not work in this case, as it would require full access to every article that is indexed by the search engine.

The results of this study would indicate that TTR was a more reliable predictor of citation count, as opposed to the GFI and DCS algorithms. This effect could be related to the tokenization process described in the method section; however, it is also possible that the GFI and DCS emphasis on word length and sentence length as an indication of complexity is less effective as a measure of intrinsic CL. A further examination of these three algorithms with a control on proper grammatical structures, would validate if this significant effect is only shown with TTR, and if GFI/DCS do not significantly impact citation count. The study also confirms that the three scientific disciplines involved varied significantly in seven out of nine cases, similar to previous research. These results show that the cognitive load generated by text complexity can contribute as an influencing factor to citation count. As the research field is growing in numbers of articles being published, it becomes more difficult to find and utilize other researchers work and an understanding the role of cognitive load on the popularity of scientific literature grows in importance.

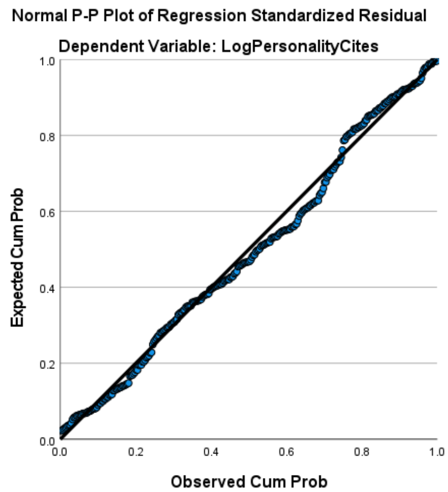
Appendix A

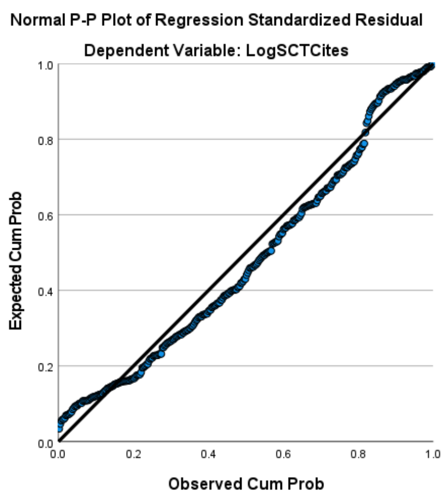

Histogram
Dependent Variable: LogPersonalityCites
Mean = 5.21E-15
Std. Dev. = 0.995
N = 296
Frequency
Regression Standardized Residual

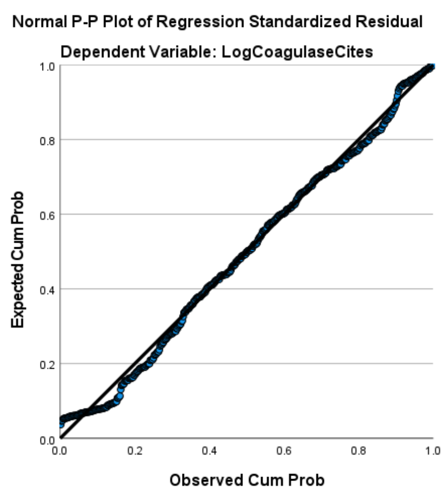

Histogram
Dependent Variable: LogSCTCites
Mean = 3.78E-15
Std. Dev. = 0.995
N = 334
Frequency
Regression Standardized Residual


Histogram
Dependent Variable: LogCoagulaseCites
Mean = 7.98E-15
Std. Dev. = 0.997
N = 454
Frequency
Regression Standardized Residual

Appendix B


Normal P-P Plot of Regression Standardized Residual
Dependent Variable: LogPersonalityCites
Expected Cum Prob
Observed Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0


Normal P-P Plot of Regression Standardized Residual
Dependent Variable: LogSCTCites
Expected Cum Prob
Observed Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0


Normal P-P Plot of Regression Standardized Residual
Dependent Variable: LogCoagulaseCites
Expected Cum Prob
Observed Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0

Appendix C

Table 3. Descriptive Statistics

		Gunning Fog Index				Dale-Chall Score				Type-Token Ratio				Citations				
	n	*Mean*	*SD*	*Min*	*Max*	*Mean*	*SD*	*Min*	*Max*	*Mean*	*SD*	*Min*	*Max*	*Median*	Q_1-Q_3	*Min*	*Max*	*No. with ze citations (%*
Personality Trait Theory	372	20.4	3.3	7.9	33.6	12.3	1.1	9.5	16.6	0.5	0.1	0.3	0.8	8.0	1.0-24.3	0	1360	76 (20.4)
Social Contract Theory	579	20.4	4.1	3.6	39.9	11.8	1.1	8.8	17.9	0.5	0.1	0.3	0.8	1.0	1.0-6.0	0	542	246 (42.3)
Coagulase Test	711	19.0	4.2	11.9	39.5	13.4	1.3	9.9	19.6	0.5	0.1	0.3	0.7	2.0	0.0-10.0	0	310	257 (36.1)

SD: standard deviation, Q_1: 1st quartile, Q_3: 3rd quartile

References

Amjad, T., Rehmat, Y., Daud, A., & Abbasi, R. A. (2019). Scientific impact of an author and role of self-citations. *Scientometrics, 122*(2), 915–932. https://doi.org/10.1007/s11192-019-03334-2

Atkinson, R. C., & Shiffrin, R. M. (1968). Human Memory: A Proposed System and its Control Processes. *Psychology of Learning and Motivation*, 89–195. https://doi.org/10.1016/s0079-7421(08)60422-3

Baddeley, A. (1984). Reading and Working Memory. *Set: Research Information for Teachers,* (1). doi:10.18296/set.1226

Baddeley, A. (2012). Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology*, 63(1), 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Besuglov, E., & Crasselt, N. (2020). The effect of readability and language choice in management accounting reports on risk-taking: an experimental study. *Journal of Business Economics,* 1. doi:10.1007/s11573-020-00980-4

Camina, E., & Güell, F. (2017). The Neuroanatomical, Neurophysiological and Psychological Basis of Memory: Current Models and Their Origins. *Frontiers in Pharmacology, 8*. doi:10.3389/fphar.2017.00438

Chall, J. S. & Dale, E. (1995). *Readability Revisited: the New Dale-Chall Readability Formula*. Brookline Books.

Cowan, N. (2010). The Magical Mystery Four. *Current Directions in Psychological Science,* 19(1), 51–57. https://doi.org/10.1177/0963721409359277

Cowan, N., Elliott, E. M., Scott Saults, J., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. Cognitive Psychology, 51(1), 42–100. https://doi.org/10.1016/j.cogpsych.2004.12.001

Debue, N., & Leemput, C. V. (2014). What does germane load mean? An empirical contribution to the cognitive load theory. *Frontiers in Psychology, 5*. doi:10.3389/fpsyg.2014.01099

Destefano, D., & Lefevre, J. (2007). Cognitive load in hypertext reading: A review. *Computers in Human Behavior, 23*(3), 1616-1641. doi:10.1016/j.chb.2005.08.012

Dewitte, S., Pandelaere, M., Briers, B., & Warlop, L. (2005). Cognitive Load has Negative After Effects on Consumer Decision Making. *SSRN Electronic Journal*. doi:10.2139/ssrn.813684

DiMascio, C. (2020). py-readability-metrics (1.3.5) [Software project]. Retrieved December 5, 2021, from https://github.com/cdimascio/py-readability-metrics

Drutman, L. (2016). Is Congress Getting Dumber, or Just More Plainspoken? Retrieved January 6, 2021, from https://sunlightfoundation.com/2012/05/21/grade-level-congress/

Garfield, E. (2006). The History and Meaning of the Journal Impact Factor. *JAMA, 295*(1), 90. https://doi.org/10.1001/jama.295.1.90

Ghanimi, H.M.A., Ab Aziz, A. & Ahmad, F. (2016). On modelling cognitive load during reading task. *Malaysian Journal of Human Factors and Ergonomics*, *1*(1).

Harzing, A. W. (2007). Publish or Perish (7.27.2949) [Computer software]. Retrieved December 5, 2021, from https://harzing.com/resources/publish-or-perish

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95.

JetBrains (2020). PyCharm Community Edition (2020.3) [Computer software]. Retrieved December 5, 2021, from https://www.jetbrains.com/pycharm/

Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Lund University, Dept. of Linguistics and Phonetics Working Papers, 53*.

Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and Instruction, 12*(1), 1-10. doi:10.1016/s0959-4752(01)00014-7

Kyle, K. (2020). lexical-diversity (0.1.1) [Software project]. Retrieved December 5, 2021, from https://github.com/kristopherkyle/lexical_diversity

Luna, K., Albuquerque, P. B., & Martín-Luengo, B. (2018). Cognitive load eliminates the effect of perceptual information on judgments of learning with sentences. *Memory & Cognition*, *47*(1), 106–116. https://doi.org/10.3758/s13421-018-0853-1

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

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*(2), 81–97. doi:10.1037/h0043158

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 58(3), 193–198. https://doi.org/10.1037/h0049234

Plavén-Sigray, P., Matheson, G. J., Schiffler, B. C., & Thompson, W. H. (2017). The readability of scientific texts is decreasing over time. *ELife, 6*. doi:10.7554/elife.27725

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422. doi:10.1037/0033-2909.124.3.372

Rayner, K., & Reichle, E. D. (2010). Models of the reading process. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(6), 787–799. doi:10.1002/wcs.68

Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B. (., & Wang, K. (2015). An Overview of Microsoft Academic Service (MAS) and Applications. *Proceedings of the 24th International Conference on World Wide Web*. doi:10.1145/2740908.2742839

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295–312. doi:10.1016/0959-4752(94)90003-5

Thompson, D. (2014). Presidential Speeches Were Once College-Level Rhetoric
Now They're for Sixth-Graders. Retrieved January 6, 2021, from https://www.theatlantic.com/politics/archive/2014/10/have-presidential-speeches-gotten-less-sophisticated-over-time/381410/

Van Merriënboer, J. J. G., & Sweller, J. (2010). Cognitive load theory in health professional education: design principles and strategies. *Medical Education, 44*(1), 85–93. doi:10.1111/j.1365-2923.2009.03498.x

Van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual. Scotts Valley, CA: CreateSpace.

White, K. (2019, December 17). *Publications Output: U.S. Trends and International Comparisons | NSF - National Science Foundation*. National Sciences Foundation. https://ncses.nsf.gov/pubs/nsb20206/

Zumbach, J., & Mohraz, M. (2008). Cognitive load in hypermedia reading comprehension: Influence of text type and linearity. *Computers in Human Behavior*, 24(3), 875–887. https://doi.org/10.1016/j.chb.2007.02.015