Sara Fagerlund

Abstract readability and online attention

—an altmetric study of research articles with Finnish affiliations

Master's thesis in Governance of Digitalization

Supervisors: Kim Holmberg & Kristina Eriksson-Backa

Faculty of Social Sciences, Business and Economics

Åbo Akademi University

Åbo 2022

ÅBO AKADEMI UNIVERSITY – Faculty of Social Sciences, Business and Economics

Abstract for Master's thesis

Subject: Governance of Digitalization	
Writer: Sara Fagerlund	
Title: Abstract readability and online attention	
Supervisor: Kim Holmberg	Supervisor: Kristina Eriksson-Backa

Abstract: Scholarly communication is developing due to social media, and research funders increasingly demand that researchers provide proof of the social impact of research. Therefore, researchers must produce articles that attract the attention of readers. The abstract can be the determining factor if research is cited or shared online. Therefore, it is beneficial to understand how readable the abstracts are and what effect readability has on online attention. Academic texts are difficult to understand, and concerns have arisen regarding accessibility. The common readability formulas, such as the Flesch reading ease, are often used to assess text difficulty, and research has mainly focused on readability and scientific impact. This study maps the readability of abstracts using new readability assessment tools and investigates the relationship between abstract readability and online attention. Lexical and syntactic complexity tools evaluated the readability of the abstracts and Altmetric attention scores were used to assess the level of online attention the research articles received. Spearman's correlation coefficient was used to analyse the relationship between readability and online attention. The data consisted of research article abstracts from 15 research areas with affiliations to Finnish universities published in 2018. The dataset for the readability analysis consisted of 9272 abstracts. From this data, 5604 articles with at least one attention score were selected for the correlation analysis. The readability analysis showed that all research areas use similarly complex language, and research areas with higher lexical complexity generally score lower in syntactic complexity and vice versa. The implication is that the studied abstracts would be too difficult to read for general audiences. The results from the correlation analysis indicate that readability has a small impact on online attention. Depending on the research area and complexity index, the effect was either positive or negative, which provides inconclusive results regarding the influence of readability on online attention. More research is needed on readability and its effect on online attention.

Keywords: abstract readability, altmetrics, Altmetric Attention Score, online attention

Date: 10.05.2022	Number of pages: 88

Table of contents

List of tables

List of figures

List of appendices

1. INTRODUCTION

1.1 Area of research

Researchers are faced with competition for the limited time and attention of readers and limited research funding, while the number of research articles produced increases (Bornmann & Mutz, 2015) and scholarly communication practices and research evaluation evolve due to digitalisation and the growing presence of social media. Producing research articles that will be read is important for the increased chances of visibility in academic contexts and online. The abstract is the first text of the research article the reader encounters while searching for information (Hartley, 2008) and can be the determining factor if the article will be read, cited, or shared online. The abstract summarises the main contents of the article and needs to convey information efficiently to the reader (Weil, 1970). Hence, there is an interest in understanding how readable the abstracts are since they are more likely to be read than the article itself and . In extension, there is an interest in understanding how abstract readability influences the impact of research.

Studies on which variables influence scientific impact are plenty (Björk & Solomon, 2012; Larivière et al., 2015; Beaudry & Larivière, 2016); however, understanding factors that drive online attention is becoming important as well. Research is increasingly discussed, shared, and disseminated on social media. For instance, scholars in bibliometrics report that social media affect their professional lives and predict it will continue to do so (Haustein et al. 2014). Online spaces create opportunities to share various scholarly outputs, and scientists can reach a wider readership. However, the amount of information available online creates information overload. Scholars have expressed difficulty finding relevant documents and experience the social media landscape challenging to navigate (Alhoori, 2019). Ease of information retrieval is important for scholars and non-scholars alike. Additionally, organisations, universities, and governments responsible for the decision-making of research funding demand evidence of impact outside of academia, such as social impact (Thelwall, 2020). For instance, Academy of Finland stresses the importance of both scientific impact and the social relevance and impact of research in their reviewing process:

> The Strategic Research Council (SRC) established within the Academy funds solution-oriented and phenomenon-driven research. The projects to be funded must have societal relevance and impact and be of a high scientific standard. Interaction with society is of key importance throughout the projects' funding periods. The scientific quality and societal relevance of projects are peer reviewed by both national and international experts (Academy of Finland, n.d.).

Research evaluation, which falls under bibliometrics and scientometrics research, has focused on analysing the number of citations articles receive and other metrics based on, for instance, productivity to reflect scientific impact (Pagell, 2014). However, they cannot account for impact outside academic contexts (Holmberg, 2015). Altmetrics are a new field in metrics research and scholarly evaluation. Altmetrics, which are alternative metrics to citations (Haustein et al., 2016), are believed to decrease the information overload in online spaces (Haustein, 2016), since altmetrics set out to measure online attention and social media mentions (Priem et al., 2010). Additionally, altmetrics are believed to open scientific discussion to the broader public, while pushing for further democratisation and transparency of science (Daraio, 2012). Studies in altmetrics have focused on various article characteristics and document properties that influence online visibility (Haustein et al., 2015; Richardson et al., 2021; Holmberg & Park, 2018). However, few have studied the effects of readability and online attention.

The growing importance of social impact and transparency of research, and open access raises questions if academic texts are too difficult to read for the wider public, especially in medicine and health-related fields (Smith et al., 2017). Furthermore, if social impact is to be used in research evaluation, the readability of research can play a role in how non-scholars engage with research. Additionally, Worrall et al. (2020) express concerns over the readability of research and the problem with misinformation (the distribution of misleading or false information, which can be deliberate or nondeliberate) across media and social media. Therefore, more readable texts are desired. Readability has been widely researched in academic contexts and various disciplines (Gazni, 2011; Lee & French, 2011; Yeung et al., 2018; Haas et al., 2018). It is generally understood that academic writing is difficult to understand and uses highly complex language (Snow, 2010; Hartley, 2008). Additionally, academic texts have become less readable over time (Plavén-Sigray et al., 2017), which decreases accessibility and contradicts the aim for openness in research.

Research on readability and scientific impact shows that research articles with more difficult abstracts receive more citations (Gazni, 2011). However, others have found that readability has no effect on scientific success (Lei & Yan, 2016) or negatively impacts citations (Didegah & Thelwall, 2013). Few studies have focused on readability and online attention. Jin et al. (2021) is the most recent study that solely focuses on readability and its influence on online attention based on a wide variety of altmetric sources. Other studies have focused on usage metrics such as the number of bookmarked and downloaded articles (Guerini et al., 2012; Chen et al., 2020). These studies indicate that readability influences the level of attention and visibility research receives online to varying degrees. However, these studies have focused on one journal or one discipline and a clear consensus on the effects of readability and online attention is yet to be formed. More research is needed that incorporates more disciplines.

Readability can be assessed using a readability formula, such as the Flesch Reading Ease, Dale-Chall, the Gunning Fog Index, and SMOG (Simple Measure of Gobbledygook), based on sentence and word counts, which result in an aggregated score that determines reading grades. The readability formulas mentioned above have been criticised for their lack of validity and predictability (Begeny & Greene, 2014; Crossley et al., 2019), and new readability tools have been developed, which use natural language processing (NLP) and machine learning that perform better at predicting readability (Benjamin, 2012; Crossley et al., 2011). New software can analyse more language features that point to specific aspects of the text that can be changed to increase readability (Graesser et al., 2011). Most studies on readability in academic contexts use the common readability formulas, or similar formulas, to assess text difficulty (Gazni, 2011; Didegah & Thelwall, 2013; Yeung et al., 2018). Few studies have used newer readability tools to evaluate the readability of academic abstracts, especially in connection to the influence of readability on scientific impact or online attention. Jin et al. (2021) and Chen et al. (2020) have used new readability and language assessment tools that focus on multiple lexical (word and vocabulary) and syntactic (grammar) features, but more research is needed to better understand how new tools can be used for the assessment of academic texts.

Article characteristics are factors individual scientists can control. Researching the factors influencing the online attention of scientific output does not only improve our understanding of altmetrics but can also help researchers write higher impact articles. Mapping the syntactic and lexical complexity of academic abstracts helps our understanding of which language features are used in academic abstracts and the level of complexity across research areas, which adds to the literature of readability.

1.2 Aim of thesis and research questions

The aim of this thesis is to map the readability of abstracts from research articles affiliated with Finnish universities published in 2018 using readability tools based on natural language processing and assess the relationship between abstract readability and the online attention of these research articles. The objective is to evaluate the lexical and syntactic complexity of different research areas and investigate if abstracts readability influences the online attention of research. The goal of this thesis is to increase knowledge on what factors influence the online attention of research and contribute to the use of readability tools based on natural language processing and add to the understanding of the level of lexical and syntactic complexity of academic language.

The research questions are:

Research question 1: What is the lexical and syntactic complexity of research article abstracts? Are there differences between research areas?

Research question 2: What is the relationship between readability and online attention? Are there differences between research areas?

1.3 Structure of thesis

First, I will present the literature review on scholarly communication, metrics research, altmetrics, and readability. Metrics research is divided into informetrics, bibliometrics, scientometrics, and webometrics. In this section, I will also discuss bibliometric indicators and criticism of metrics research. In the next part, altmetrics will be addressed and is divided into a discussion on the definition of altmetrics, what the altmetric indicators and altmetric sources are, the providers of altmetrics, factors influencing online attention, and lastly, a discussion on the advantages and disadvantages of altmetrics. The chapter on readability begins with a presentation of what readability is, and then a section follows on the most common readability formulas and new developments in readability. Lastly, I will discuss criticism aimed at readability and readability programs. Then a review of research on academic language and readability will follow. I will first discuss the readability of academic language in general and the features of the academic abstract. Earlier research on readability and scientific impact will follow, with the last section focusing on earlier research on readability and the online visibility of research. In the data collection and method section, I will discuss the data selection process and methods used. Then I will present and discuss the results of the readability analysis and correlation analysis. Lastly, I will explain the limitations of the study and offer recommendations for future research.

2. LITERATURE REVIEW

2.1 Scholarly communication

Borgman and Furner (2005:13) defines scholarly communication as "how scholars […] use and disseminate information through formal and informal channels". Scholarly communication starts with a research idea, a written research paper, followed by the peer review process, resulting in a published paper discussed, cited, and referenced by other researchers (Holmberg, 2015). Weimer and Andrew (2013) describe the life cycle of scholarly communication, both formal and informal:

> The research lifecycle begins with an idea to pursue, followed by data collection, and data analysis, and continues with creating a story, or context for the analysis. The product of that analysis could be shared in the form of a book or article, blog, illustration, presentation, or other communication channel. These resources, or information vehicles, then provide an opportunity for the scholarly community to engage in conversation, debate, and further study on the topic at hand. The outcome of further study starts the cycle anew. (Weimer & Andrew, 2013: 217)

The peer-reviewed journal article is still regarded as the most crucial channel of scholarly communication (Weimer & Andrew, 2013; Haustein et al., 2015). Formal scholarly communication refers to published research, such as the journal article, whereas informal scholarly communication is conversations among colleagues, emails, phone calls, and social media.

Scholarly communication is a complex ecosystem of dependent and competing relationships between stakeholders (Jubb, 2013). Jubb (2013) summarises the main stakeholders as researchers, universities and other research institutions, research funders, librarians, publishers, and learned societies. These stakeholders have their own needs and roles in the scholarly communication process. Researchers need efficient publication and dissemination practices to increase their chances of publishing their research in high-quality journals and get credit for their work (Jubb, 2013). They are also readers and producers of research, who require easy and free access to information. Universities and research institutions need to maximise their research performance and decrease expenses. Research funders must ensure that the research they fund produces high impact and is widely accessible. The role of librarians is to maximise the availability of high-impact and high-quality publications and provide the necessary services to researchers. Publishers offer services for the publication and dissemination of research. These stakeholders all seek to minimise the cost of their practices (Jubb, 2013).

Informal scholarly communication and social media

Scholarly communication relies on scholars finding relevant information. The massive amount of information available has long been an issue in science, with citation indexing as the first step towards easier information retrieval (Haustein et al., 2015). New challenges regarding big data, which are data too large and varied for traditional data processing software, have arisen due to social media and digital technologies.

Social media has become a part of informal scholarly communication, and platforms such as Twitter are increasingly used. Haustein et al. (2015) write that between 70 and 80% of scholars use social media. Holmberg and Thelwall (2014) found disciplinary differences in Twitter use. Researchers in humanities use Twitter for discussions, whereas biochemists retweet more, and researchers in economic disciplines share more links (Holmberg & Thelwall, 2014). Mendeley, a social reference manager, also exhibits differences among disciplines. Zahedi and van Eck (2018) found that social sciences and humanities have the most active readership, which was surprising since these disciplines have lower citation counts than other fields.

Shehata et al. (2015) write that scholars have become more inclined towards informal scholarly communication practices. Scholars are increasingly using informal channels to publish and disseminate research or parts of their research; however, peer-reviewed articles are still held in higher regard concerning the quality of research (Shehata et al., 2015). Additionally, scholars who do not publish research in informal channels still use them to promote their scholarly profile when sharing their research with a wider audience (Shehata et al., 2015). Shehata et al. (2015) add that the current peer review process must be improved since many scholars believe it hinders important research articles from being published.

Open access

Houghton et al. (2004) write that the system through which scholars communicate is important in the facilitation of research dissemination, therefore, ease of access is essential. The push for open access publishing has challenged the status quo of the peer-reviewed journal article as the most influential aspect of scholarly communication (Weimer & Andrew, 2013). Scholars do research for free, and the peer review process is free, but the journals come at a cost. Open access is the answer to this disparity and gained support partly due to the internet enabling easy and free information sharing (Yiotis, 2005). Scholars wanted to take back ownership of their work and provide free and easy access of research for everyone (Yiotis, 2005). Morrison explains the fundamental concepts of open access:

> Open access is scholarly literature that is digital, online, free to read and free of most copyright and licensing restrictions. Open access can be green, when authors self-archive their work for open access, or gold, when the publisher makes the work open access. Open access can be gratis (free to read) or libre (free to read and reuse). Open access can apply to the works themselves, or to the process of making works open access. (Morrison, 2009: 133)

The growth of open access is illustrated by the number of open access journals indexed by DOAJ (Directory of Open Access Journals). DOAJ contains over 17.000 open access journals (DOAJ, n.d.). Open access is seen as an important step in the democratization of scholarly communication and enables faster publishing (Morrison, 2009). However, Osborne (2015) argues that open access is not inherently accessible, and the cost of research is not the main issue. Osborne (2015) states that accessibility is achieved through well-written research and increasing the number of editors and referees. Green (2019), on the other hand, says it is the publishing process and cost that are the main problems, not open access itself.

2.2 Metrics research

Scholarly communication is the exchange of ideas, both in formal and informal channels and may result in articles being cited and referenced. This exchange of ideas and influence needs to be captured and evaluated to help facilitate decision-making about all facets of academia, from research funding to job offers and university rankings. In this part, I will present the key areas of metrics research that lays the foundation for altmetrics. First, I will present informetrics, which is the broadest area. Then I will present bibliometrics, scientometrics, webometrics, and the advantages and disadvantages of metrics research and scholarly evaluation tools.

Informetrics

Informetrics is defined as the quantitative measure of information. Tague-Sutcliffe (1992:1) describes informetrics as “the study of quantitative aspects of information in any form, not just records or bibliographies, and in any social group”. Hood and Wilson (2001) write that there is confusion surrounding the terminology of informetrics and bibliometrics since these terms are used interchangeably to refer to similar concepts. They explain the definition of informetrics and the background behind the term:

> The most recent metric term, ‘informetrics’, comes from the German term ‘informetrie’ and was first proposed in 1979 by Nacke to cover that part of information science dealing with the measurement of information phenomena and the application of mathematical method to the discipline’s problems, to bibliometrics and parts of information retrieval theory (Hood & Wilson, 2001: 294).

As Tague-Sutcliffe (1992) further explains, informetrics measure informal, formal, spoken, and written communication and is focused on areas outside the scope of bibliometrics. Hood and Wilson (2001) reiterate that informetrics has a broader scope than bibliometrics. Informetrics, bibliometrics, scientometrics, webometrics, and altmetrics are part of metrics research (Figure 1), measuring quantifiable units of information in different domains and employing various methods to assess the specific needs of each subfield. Informetrics is now frequently used as an umbrella term for bibliometrics, scientometrics, webometrics, and altmetrics (Egghe, 2005; Holmberg, 2015).

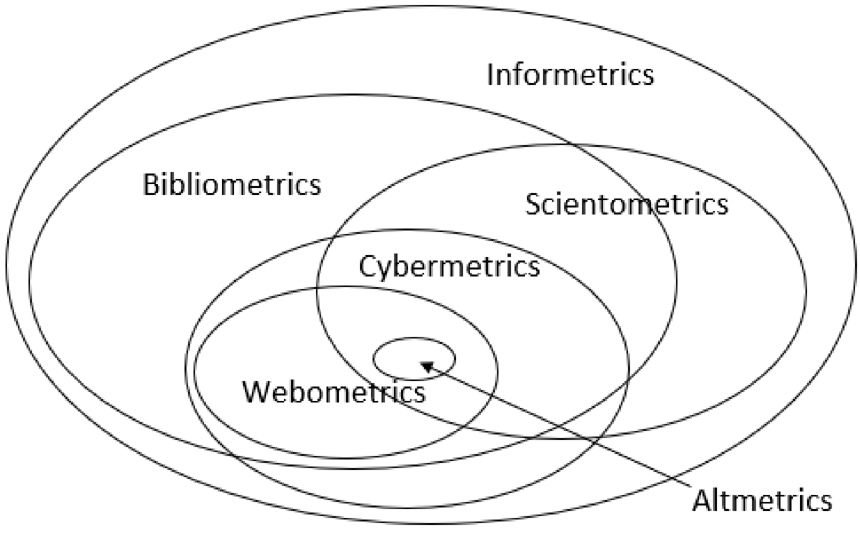

***Figure 1**. Visualisation of metrics research (Holmberg, 2015: 15)*

Bibliometrics

Bibliometrics has a long tradition in library and information science. The term bibliometrics was first mentioned by Pritchard (1969) and defined as "the application of mathematics and statistical methods to books and other media of communication". The definition of bibliometrics, as described by Pritchard (1969), has been criticised due to its vagueness (Broadus, 1987). Broadus (1987) presents an overview of the development towards a definition of bibliometrics and concludes that, despite being difficult to define, bibliometrics is the logical definition of the field of research dealing with units of publications and bibliographic citations. Tague-Sutcliffe (1992) defines bibliometrics as:

> The study of the quantitative aspects of the production, dissemination, and use of recorded information. It develops mathematical models and measures for these processes and then uses the models and measures for prediction and decision making (Tague-Sutcliffe, 1992: 1).

Although bibliometrics can measure any recorded document, the field is heavily focused on measuring science and scientific output. Bibliometrics as a scientific field arose after Eugene Garfield developed the Science Citation Index (SCI) in the 1960s (Thelwall, 2008). The Social Sciences Citation Index (SSI) and Arts and Humanities Citation Index emerged later, and these databases (now part of Web of Science) enable more manageable collection of statistics regarding citations and scientific impact (Thelwall, 2008). Garfield (1973) believed the frequency of citations could explain scientific quality. Citations are still the most used metric in impact and quality assessment. Bibliometrics are often highly skewed, with few journals and authors receiving the most impact and citations (Tague-Sutcliffe, 1992; Andrés, 2009; Holmberg, 2015)

There are two types of bibliometrics: evaluative bibliometrics and relational bibliometrics. Evaluative bibliometrics measure scholarly impact and is usually involved in research funding and research policy (Thelwall, 2008). Relational bibliometrics focus on relationships in research, investigating emerging research fields, and discovering patterns of research collaboration (Thelwall, 2008). Bibliometrics measure document properties using word frequency analysis, co-word analysis, citation analysis, and counting the number of documents produced by one or more authors, countries, or institutions (Holmberg, 2015). Borgman and Furner (2005) state that bibliometric methods can only be applied to written scholarly communication occurring in formal discussion channels.

Scientometrics

Bibliometric research focusing on science and scholarly communication falls under scientometric research (Holmberg, 2015). Nalimov was the first to define the concept of scientometrics in 1971, according to Mingers and Leydesdorff (2015). Hood and Wilson (2001: 293) write, “much of scientometrics is indistinguishable from bibliometrics, and much bibliometric research is published in the journal, *Scientometrics*”. The main difference is scope. Bibliometric research focus on any recorded information, such as scientific papers, and other types of media, whereas scientometrics are concentrated on scientific and technological contexts. Tague-Sutcliffe (1992) provides the following definition of scientometrics:

> Scientometrics study the quantitative aspects of science as a discipline or economic activity. It is part of the sociology of science and has application to science policymaking. It involves quantitative studies of scientific activities, including, among others, publication, and so overlaps bibliometrics to some extent. (Tague-Sutcliffe, 1992:1).

Bornmann and Leydesdorff (2014) write that scientometrics is changing due to new technological developments such as social media and web 2.0 but has established itself as an important factor in decision making regarding job offers, research funding, and research policy.

Webometrics

Webometrics is closely linked to bibliometrics and scientometrics. Similar methods of analysis are used, but, as the name suggests, webometrics aim to analyse web content (Björneborn & Ingwersen, 2001). Thelwall (2008) lists webometrics such as link analysis, web citation analysis, and search engine evaluation. According to Almind and Ingwersen (1997), web pages connected by hyperlinks are the citations of the web. Analysis of inlinks and outlinks are at the core of webometrics (Björneborn & Ingwersen, 2004). Ingwersen (1998) developed the Web Impact Factor (Web-IF) to calculate web engine performance. However, the dynamic nature of the web, lack of metadata, and lack of quality control due to the varied userbase complicate the use of bibliometric methods and Web-IF (Björneborn & Ingwersen, 2001). Bar-Ilan (2001) states that data collection on the web is unsatisfactory since quality and validity cannot be assured.

Bibliometric indicators

Bibliometrics evaluate the scientific impact and quality of research for predictions and decision-making. Publication analysis measures the productivity of an individual author, faculty, or university by counting the number of articles published. Citation analysis counts the number of citations a journal, faculty, university, or author has received. Metrics such as the H-index or Journal Impact Factor (JIF), based on the number of citations, assess the quality of individual authors or the quality of journals. Web of Science, Scopus, and Google Scholar are the most used databases in bibliometric research. Pagell (2014) lists the bibliometrics, measurements, and sources used to rank universities and institutions and evaluate scientific impact and the quality of research in Table 1.

Metric	Measurement	Sources
Publication	Number of articles Number of pages	Web of Science Scopus Google Scholar Individual databases and websites
Citations	Number per article Number per faculty Number per university Highly cited papers	Web of Science & Essential Science Indicators Scopus Google Scholar Individual databases (Science direct, EBSCO. JStor, Proquest) Scholarly websites (Repec, AMC Portal
H-Index	The number of papers with citation numbers higher or equal to the number of citations (Hirsch, 2005)	Web of Science Scopus Individual calculations
Journal Quality	Journal Impact Factor Eigenfactor SNIP SJR	Journal Citation Reports Eigenfactor.org SCImago Leiden

***Table 1**. Bibliometrics used in quality assessment and university rankings (Pagell, 2014)*

Citations link and give research credibility and value (Holmberg, 2015). However, simply counting the number of citations is not always sufficient. The Hirsch index (H-index), developed by Hirsch (2005), measure the impact and scientific output of scientist in the same discipline. Hirsch (2005) explains, "a scientist has index h if h of his or her Np papers have at least h citations each and the other (Np – h) papers have ≤h citations each". Pagell (2014) explains how the h-index works and varies across databases.

> If an author has 44 papers in SCOPUS with 920 citations and the 16th paper has 16 citations the H-Index is 16; if the same author has 36 papers in WOS with 591 cites and the 13th paper has 13 citations, the H-Index in WOS is 13. That same author created an author ID in Google Scholar, which tracks articles and citations. The author has 65 publications, 1921 citations and the 21st article has 21 citations for an H-index of 21 (Pagell, 2014: 143).

The H-index measures the quality and impact of individual authors, whereas journal quality is measured using different approaches. The Journal Impact Factor (JIF), developed by Garfield and Sher (1963), is perhaps the most well-known, and the impact factors are published in *Journal Citation Reports* by the Institute for Scientific Information (ISI). Journal impact measures how many citations a journal has received in a two-year period.

Criticism of metrics research

Evaluating the quality and impact of research is challenging. Although many different methods exist, they are often deemed inadequate. Seglen (1997) writes that science should be reviewed by specialists in the field in question, which is rarely the case, and thus, committees use citation counts and impact factors for decision making. Seglen (1997) argues that these are crude measurements of quality. Especially the journal impact factor has been heavily criticised. Leydesdorff et al. (2019) state that the original purpose of the journal impact factor was as a tool for library portfolio management. Still, it has been misused to evaluate the quality of individual papers. Leydesdorff et al. (2019) further add that an average citation rate over two or five years fails to indicate the quality of a journal. Seglen (1997) also states that impact factors are used for marketing purposes, not quality assessment.

The San Francisco Declaration of Research Assessment (DORA) was created in response to the growing dissatisfaction with the current metrics used for research evaluation. DORA also aims heavy criticism to the misuse of the Journal Impact Factor in research assessment. It has outlined guidelines for best practices for funding agencies, institutions, publishers, metrics providers, and researchers to combat unfair evaluation and misuse of metrics. The main criticisms of the Journal Impact Factor are:

- citation distributions within journals are highly skewed
- the properties of the Journal Impact Factor are field-specific: it is a composite of multiple, highly diverse article types, including primary research papers and reviews
- Journal Impact Factors can be manipulated (or “gamed”) by editorial policy
- data used to calculate the Journal Impact Factors are neither transparent nor openly available to the public (DORA, n.d.)

The Leiden Manifesto by Hicks et al. (2015) is another response to the misuse and misapplication of the H-index, JIF and data-driven quality assessment in research evaluation. The manifesto offers ten principles to abide by in research evaluation to provide a fair and transparent evaluation process.

Citations are the most important metric for quality and impact assessment. However, there are several reasons why articles are cited. An article may be cited if the researcher wishes to highlight errors or critique said article (Bloch & Walter, 2001), which does not indicate that the cited article is of high quality. Metrics based on citations such as H-index have also been criticised. As demonstrated by Pagell (2014), the H-index varies across databases. Bar-Ilan (2008) writes that if the H-index is used, awareness of the difference in disciplinary coverage in Web of Science, Scopus, and Google Scholar is essential. Bornmann and Daniel (2009) stress the importance of comparing scientists and researchers in the same discipline and with similar academic careers if the H-Index is used. Bornmann and Daniel (2009) conclude that comparing research based on numbers fails to understand scientific impact fully. Lastly, traditional bibliometric indicators only account for scientific impact, but scholarly communication is taking place in social media and data generated in these places need to be captured and measured to broaden the ways research is evaluated.

2.3 Altmetrics

The presence of social media and increasing demands for evidence of both scientific and social impact from funding bodies have affected bibliometrics and scientometrics. New ways to measure scientific output are needed to meet these demands and capture the vast amount of data generated by discussions and scholarly communication online. Shema et al. (2014) write, “altmetrics, short for alternative metrics, is a term to describe web-based metrics for the impact of scholarly material, with an emphasis on social media outlets as sources of data”. Altmetrics can be regarded as an extension of webometrics. Priem et al. (2010) write that altmetrics measure and capture scientific output in online spaces and evaluate social media acts by combining traditional bibliometric tools with new alternative metrics. Both scholars and non-scholars use social media and scholars publish output such as code and parts of research online and not only the finished research paper. Thus, analysing the online visibility and online attention can point towards the broader impact of research, such as social impact (Priem et al. 2012) and capture various scholarly output.

Additionally, altmetrics provide answers to the criticism of traditional bibliometric indicators (Priem et al., 2010; Daraio, 2021). Priem et al. (2010) argue that the journal impact factor lacks transparency and propose altmetrics are more open and transparent due to openness of data. Daraio (2021) writes that altmetrics respond to the need for democratisation of research evaluation since new online platforms allow for broader discussion of scientific output outside of academia.

Altmetrics are a relatively new field of study. Factors influencing online attention, what altmetrics capture, and how to best use altmetrics in research evaluation are still investigated. In the next section, I will present key concepts of altmetrics, such as social impact, altmetric indicators, and altmetric providers. Then, I will introduce theories and research on factors influencing altmetrics. Lastly, I will discuss the advantages and disadvantages of altmetrics.

Social impact

The main goal of altmetrics is to measure the social or societal impact of research by analysing the mentions of scientific output in online spaces (Priem et al., 2012). Although the goal is to measure social impact, research papers rarely use the term social impact. Instead, definitions such as online visibility, online attention, altmetric events, or altmetric mentions are used. Furthermore, social impact and societal impact are used interchangeably in altmetric literature (Pulido et al., 2018). I will continue to use social impact to refer to both in this thesis.

There are no agreed-upon methods that identify and capture the social impact of research (Holmberg et al., 2019a) and no agreed-upon definition of what social impact is. Social impact in the context of altmetrics can be understood as "measuring the social, cultural, environmental and economic returns from publicly funded research" (Bornmann, 2012). Social impact in the broadest sense can also be seen as the influence research has on society and the opposite of scientific impact.

Providing evidence of social impact is difficult since there are no agreed-upon methods and definitions. However, online attention can be seen as a reflection of social impact to some extent (Priem et al., 2010; Holmberg et al., 2019a). Holmberg (2015) considers visibility, influence, attention, and engagement in social media as indicators of various levels of impact. Haustein (2016), on the other hand, claims that social media does not reflect the wider social impact, not until the meaning of altmetric indicators is adequately understood. Pulido et al. (2018) argue that social media can be a tool to gain insight into the social impact of research by analysing the types of messages posted about research, not the number of messages. Bornmann (2015) suggests that the low correlation between online attention and citations points to altmetrics as a broader impact measurement of research.

As illustrated by the Academy of Finland, organisations call for evidence of various impacts. Publishers have adopted altmetric scores to give scholars insight into the level of attention their research has received. However, altmetrics are yet to be used in formal research evaluation (Thelwall, 2020). Providers of altmetric data suggest researchers use altmetrics in their grant applications if the scholar has little research background and has accumulated few citations (Altmetric.com, n.d.c). Onyancha (2019) investigates if altmetrics can be used in assessing research excellence and states that highly cited papers do not have higher online attention. Onyancha (2019) concludes that clear guidelines and frameworks are necessary if altmetrics are used in research assessment. Additionally, it is generally agreed that altmetrics cannot and should not replace citations in quality assessment and research evaluation. Altmetrics are a complement to traditional metrics (Haustein et al., 2015; Holmberg & Park, 2018; Bornmann, 2014).

Proving that research has influenced society is difficult but counting the number of online mentions provides some evidence of influence. However, a better understanding of altmetric data and indicators is needed (Haustein, 2016). Lastly, Sugimoto (2015) adds that the use of impact is a hyperbole:

> The term impact connotes far greater engagement and transformative effect than is currently justifiable with altmetric data. A more persuasive claim is that what is captured are metrics of attention of a scholarly object—the nature of this attention is something much more complex and far less understood (Sugimoto, 2015).

Altmetric indicators

There are several sources of altmetrics, referred to as altmetric indicators, that measure the online attention of research. Blogs, microblogs, social networking sites, social reference managers, and Wikipedia are among the more popular groups of indicators (Holmberg, 2015; Haustein, 2016). Other altmetric indicators are mainstream media, public policy documents, and patents. However, not every altmetric indicator reflects the same level of impact. According to Holmberg (2015), higher coverage reflects lesser impact. This is evident on Altmetric.com and their ranked list of altmetric indicators as presented in Table 2. Tweets and Facebook wall posts have the lowest altmetric score (0.25 per mention), while tweets have the highest coverage of social media indicators (Thelwall et al. 2013). News articles and blogs have the highest scores (8 and 5, respectively) on Altmetric.com, while they have low coverage (Thelwall et al., 2013). More information on the Altmetric Attention Score can be found in the Data collection and method chapter (p. 44).

Altmetric source	Score
News	8
Blog	5
Policy document (per source)	3
Patent	3
Wikipedia	3
Peer review (Publons, Pubpeer)	1
Weibo (not trackable since 2015, but historical data kept)	1
Google+ (not trackable since 2019, but historical data kept)	1
F100	1
Syllabi (Open Syllabus)	1
LinkedIn (not trackable since 2014, but historical data kept	0.5

Twitter (tweets and retweets	0.25
Facebook (only a curated list of public Pages)	0.25
Pinterest (not trackable since 2013, but historical data kept	0.25
Q&A (Stack Exchan)	0.25
Youtube	0.25
Number of Mendeley readers	0
Number of Dimensions and Web of Science citations	0

Table 2. *The weighted ranking of altmetric sources of the Altmetric Attention Score (Altmetric.com)*

Social media platforms can be divided into social networking sites such as Facebook, and microblogging sites (Haustein, 2016). Twitter is the most popular microblogging site. Social reference managers are also a part of social media (Haustein, 2016, Holmberg, 2015). Altmetric data from Twitter consists of tweets and retweets, whereas Facebook data is collected from public pages. Thelwall (2020) writes it is inconclusive how well Facebook can be used as an altmetric source due to low coverage. Twitter, however, is well studied and has high coverage of altmetric mentions. Research has focused on factors affecting the popularity of research on Twitter (Zhang & Wang, 2021), whether tweets correlate to citations (Eysenbach, 2011), and analysing disciplinary differences in twitter use (Holmberg & Thelwall, 2014).

Social reference managers are used to collect and organise research articles and connect with groups. There are several reference managers such as Zotero, CiteULike, and Mendeley. Mendeley is the most popular and researched in altmetrics (Chen et al., 2018; Thelwall, 2018; Zahedi & Van Eck, 2018). It is debated whether social reference managers should be included as altmetric indicators if the goal is to measure impact outside academia. Chen et al. (2018) discovered that mostly researchers, students, and professors use social reference managers. Zahedi and Van Eck (2018) further state differences in usage among professors and students in different disciplines. Bar-Ilan et al. (2019) state that social reference managers do not point towards social impact, but to a wider scholarly context and may still be an essential altmetric indicator. Thelwall (2018) writes that Mendeley readerships counts may predict scientific impact. Altmetric.com does not include Mendeley readership counts in their score; it is accessed separately.

Wikipedia was created in 2001 and is the world's largest online encyclopaedia. Anyone can contribute and edit articles. However, the information published on the site must be verified against reliable sources (Wikipedia, n.d.). A citation on Wikipedia is considered to have a higher impact than other metrics and may point to the social or educational impact of research (Pooladian & Borrego, 2017). The higher impact of Wikipedia is also reflected in the Altmetric Attention Score. A citation on Wikipedia is worth 3 points. The disciplines with the most Wikipedia citations are medicine, biochemistry, molecular biology, and agricultural and biological sciences (Arroyo-Machado et al., 2020). However, the coverage of Wikipedia is low. For instance, only 3% of library and information studies articles published between 2001 and 2010 have received a citation on Wikipedia (Pooladian & Borrego, 2017).

Blogs range from personal to corporate blogs. The disadvantage of blogs is the difficulty of collecting data. While social media platforms such as Twitter and Facebook are available on one specific site, blogs can be hosted by private domains or platforms designed specifically for blogs (Holmberg, 2015). Ortega (2019) writes that Altmetric.com has the highest blog coverage of the altmetric providers. Altmetric.com collects data from over 9.000 academic and non-academic blogs, enabling easier data collection. An article mentioned in a blog post receives 5 points by Altmetric.com but mentions in blogs are rare (Thelwall, 2020). Shema et al. (2012) found that blogs favour articles from high impact journals, and the most covered disciplines were from life and behavioural sciences.

Other altmetric indicators show different types of impact besides social and scientific impact. Patents may indicate the commercial impact of research, and grey literature can reveal the governmental influence of research (Thelwall, 2020). However, these indicators and news media are rarely the focus of research in altmetrics. Ultimately, Sud and Thelwall (2013) emphasise the importance of context in altmetrics, stating “altmetric indicators may be relevant and valid in certain contexts but not in others” and reasons for sharing research are different on different platforms.

Altmetric providers

PlumX and Altmetric.com are two of the biggest providers of altmetric data. PlumX was founded in 2012 by Plum Analytics and acquired by Elsevier in 2017. PlumX divides its metric into five categories: citations, usage, captures, mentions, and social media (Plum Analytics, n.d.a). PlumX captures metrics for all types of research output, called artefacts, and collects metrics of 67 different types of output (Plum Analytics, n.d.b).

Altmetric.com was created by Euan Adie in 2011 and supported by Digital Science (Altmetric.com, n.d.c). Their database covers over 35 million research outputs and over 191 mentions as of 2022. Altmetric.com tracks various output types such as policy documents, news, blogs, social media, Wikipedia, patents, and more (Altmetric.com, n.d.b). These sources make up what Altmetric.com calls the donut, which visualises the aggregated Altmetric Attention Score (Altmetric.com, n.d.b). Citations and Mendeley readership counts are not included in the Altmetric Attention Score but can be accessed through the database Altmetric explorer (Altmetric.com, n.d.a).

Differences between altmetric providers have been researched. Ortega (2018) states that Altmetric.com captures more online mentions than PlumX. The coverage of Twitter, news, and blogs is higher on Altmetric.com. However, Wikipedia citations and Mendeley readership counts are higher on PlumX (Ortega, 2018). Ortega (2018) concludes that despite Altmetric.com performing better, it is helpful to collect metrics from both providers due to the varying coverage. Bar-Ilan et al. (2019) confirm discrepancies between coverage on Altmetric.com and PlumX, although these differences have decreased over the years. Bar-Ilan et al. (2019) support the use of multiple altmetric providers and point to the different methodologies and metrics used by the providers as reasons for differences in coverage of altmetric sources. Karmakar et al. (2021) provide similar findings and add that PlumX tracks more research output and altmetric sources than Altmetric.com.

Factors influencing the online attention of research

There are several theories on citing behaviour, such as the normative theory, constructivist theory, and the Matthew effect. However, theories on why specific research is shared online have yet to be conceptualised in altmetrics. Haustein et al. (2016) propose a few theories based on the theories mentioned above and present social theories such as social capital, attention economics, and impression management. Research in altmetrics has focused on users engaging in sharing, discussing, and dissemination of scientific output online and document properties and characteristics of research influencing online visibility. Citation theories will be presented first, then the social theories. Lastly, studies on factors influencing online visibility will be presented.

Haustein et al. (2016) discuss whether the normative and social constructivist theories apply to altmetrics. The Normative theory states that scientists follow certain social norms in their citing behaviour (Haustein et al., 2016). Merton (1973) explains citing behaviour as crediting the researcher or scientist by citing their work. Bornmann explains that citations represent “intellectual or cognitive influence on scientific work” (Bornmann, 2006: 48). The social constructivist perspective argues that the cognitive content of articles has little impact on their reception. This perspective on citations claims that research must advocate for itself and its result to be valued by the rest of the scientific community (Bornmann, 2006), the so-called persuasion hypothesis (Haustein et al., 2016). A part of these theories is the Matthew effect. Merton (1968) wrote that the Matthew effect focuses on reward systems in science. Merton (1968) found that well-known and established scientists receive more credit and citations in collaborations than their lesser-known counterparts. If a researcher is already established and recognised, they have an easier time to be recognised again and get access to resources and acclaim. The Matthew effect influences all areas of the scientific institution, from citations to research funding (Merton, 1968). Haustein et al. (2016) write that the Matthew Effect is a promising theory for explaining Twitter behaviour in scholarly contexts.

Social theories of online behaviour are social capital, impression management, and attention economics. Social capital is a source of power from being in a social network where each actor benefits from the relationship (Haustein et al., 2016). Erving Goffman, who first explained impression management, claimed that people have specific roles and perform in social situations to avoid shame and embarrassment in order to present desirable information about themselves to others (Segre, 2014; Haustein et al., 2016). Impression management has been linked to tweeting behaviour, since people form an opinion of the tweeter based on the contents of tweets (Haustein et al., 2016). Lastly, attention economics “considers the costs and benefits of finding useful information” (Haustein et al., 2016: 10). Information overload is a problem in scientific contexts, and there is competition over people’s attention which is limited and thus highly valued (Haustein et al., 2016).

Factors influencing online attention have also been studied. Holmberg and Park (2018) discovered that only a minority of articles receive the most online attention and open access is the main driving factor. However, there are disciplinary and platform differences in the level of attention open access articles receive (Holmberg et al., 2020). Online attention is also driven by collaboration. Holmberg et al. (2019b) found that international collaboration influenced the number of tweets. Haustein et al. (2015) support the findings that collaboration positively influences social media attention. The reason for increased attention may be due to international visits creating new networks that drive attention to more research internationally (Holmberg et al., 2019b). Another factor that drives online attention is the number of references (Haustein et al., 2015). Richardson et al. (2021) also found that the number of references increased the online attention of research. Contrastingly, Zhang and Wang (2021) write that influential accounts, timing, and accessibility influence attention on Twitter, not document properties. Others have expressed theories that certain articles are shared online due to their funny titles (Sugimoto, 2015).

Research on citing behaviour focus on participants whose intentions and citers as a group are chiefly uniform, whereas investigating behaviour in information sharing on social media is exacerbated by the variety of users and their reasons for engaging with research online. Efforts have been made to map the behaviour of users who engage in the dissemination of research on social media. Differences in demographics were found in the use of social reference managers. Chen et al. (2018) found that Mendeley users were younger and used by all genders. In contrast, Zotero users exhibited more engagement in social media and had a background in social sciences and humanities (Chen et al., 2018). Holmberg and Thelwall (2014) found that biochemists, humanities, and astrophysics use Twitter the most, while economists use twitter the least. Vainio and Holmberg (2017) write that research is shared by Twitter users who define themselves by their expertise rather than their personal interests, which implies that most people who share research online are scholars and other experts.

Advantages of altmetrics

One of the main advantages of altmetrics is the possibility of capturing the immediate social impact of research (Garcia-Villar, 2021). Although it is debated whether altmetrics capture social impact, it quickly captures the online attention research receives on multiple platforms. Fraumann (2018) writes that this opens possibilities for new researchers and established researchers alike. New researchers can garner attention in the scientific community early in their career if their research is visible online, while established researchers can receive online attention without being active on social media themselves (Fraumann, 2018).

Additionally, altmetrics measure broader output types and impact outside scientific contexts (Thelwall, 2020). Bornmann (2014) writes, “altmetrics could reveal impact which traditional indicators have hitherto been unable to reveal”, such as policy change and clinical practice. Altmetrics capture output such as patents and code previously unmeasured by traditional bibliometrics (Priem et al., 2010). Furthermore, altmetrics provide multiple indicators for specific research assessment contexts and needs (Thelwall 2020).

Altmetrics is believed to represent disciplines such as social sciences and humanities better than bibliometrics (Costas et al., 2015). Humanities mainly publish research in books, while bibliometrics usually focus on citations in journals, resulting in a lack of coverage of humanities research (Hammarfelt, 2014). The journal impact factor is criticised for lacking transparency since it is unclear how it is calculated (Priem et al., 2010). Data is freely accessed through web application programming interfacing (API), enabling a more straightforward analysis of data (Bornmann, 2014). Lastly, discussions surrounding research moving to social media generates accessibility for the wider public.

Disadvantages of altmetrics

Although altmetrics presents many opportunities, several challenges and disadvantages must be considered. Haustein (2016) lists three main challenges of altmetrics: heterogeneity of data, data quality, and dependencies on platforms and technology. Heterogeneity of data refers to the diversity of online events, which complicates data collection and understanding of what altmetric data entail (Haustein 2016). Data quality declines due to the nature of social media, resulting in lack of accuracy, consistency, and replicability of altmetric data (Haustein, 2016). Altmetrics are shaped by social media platforms, data providers, and the technology available to track altmetric events such as digital object identifier (DOI) and APIs (Haustein 2016). DOI has become more common; however, if DOI is missing, the research will not be captured by altmetrics. Furthermore, DOI is only used for articles, which results in other scientific output being left out of the evaluation process (Haustein 2016). Altmetrics claim to capture broader output types, but the reliance on DOI limits altmetric research to journal articles. Altmetrics are only a snapshot of a specific moment in time due to the dynamic nature of data (Holmberg et al., 2019a). It is important to note when the data was gathered, since social media posts can be removed, and social media platforms cease to exist.

The lack of a theoretical framework is another criticism of altmetrics (Haustein 2016). As discussed earlier, one of the most essential aspects of altmetrics lacks a clear definition, social impact. It is difficult to measure impact due to the nature of social media (Holmberg et al., 2019a), and it is, therefore, challenging to conceptualise altmetrics. Reasons for posting online differ, and the habits and behaviour of users vary depending on the platforms. Theories explaining why certain research receives more online attention are needed. Haustein et al. (2016) have made efforts to incorporate theories from social constructivism, attention economics, and citation theories to altmetrics (see pp. 24-25).

Altmetrics can visualise the level of online visibility research receives, but it fails to indicate the quality of research (Fraumann, 2018; Thelwall, 2020). Research can be shared and garner discussion for different reasons that are not always positive, and qualitative analysis is needed to assess altmetrics (Fraumann, 2018; Holmberg, 2015). The importance of certain altmetric indicators may also be overstated by altmetric providers since they are owned by private and for-profit businesses (Haustein, 2016).

Lastly, altmetrics may be prone to manipulation. Altmetrics are becoming an important aspect of scholarly communication. If altmetrics are to become a part of scholarly evaluation, the risk of manipulation increases, and bots may be used to manipulate the level of engagement (Holmberg, 2015).

2.4. Readability

Factors influencing altmetrics have ranged from social theories to document properties. Language features and readability are also aspects of interest in scholarly evaluation. Readability measures the difficulty of a written text. If the text is not understood by the intended reader, it is not readable (Klare, 1963). The linguistic aspects that influence readability are lexical (words and vocabulary) and syntactic (grammar) features. The layout and typeface of the text are excluded from readability but are linked to legibility (the deciphering of symbols related to the graphic and stylistic aspects of the text), according to DuBay (2004). However, legibility may influence how the reader experiences the text (Sheedy et al., 2005).

Janan and Wray (2014) identify three main elements of readability: comprehension, reading speed, and motivation. Firstly, Perfetti (2007: 357) writes, "in reading, the singular recurring cognitive activity is the identification of words" and suggests text comprehension is linked to the recognition of words. Perfetti (2007) states that a broader vocabulary determines how well the reader understands the text. Secondly, readability is linked to the optimal speed of reading (Janan & Wray, 2014). Less readable texts require more attention and are more time-consuming. Just and Carpenter (1980) found that the average reading speed is 200 words per minute if the text is read to be comprehended. Lastly, motivation, personal experiences, and previous knowledge are factors to consider in readability. Dale & Chall (1948) stress that texts will take longer to read if the reader is uninterested in the topic. Just and Carpenter (1980: 351) conclude that "a well-written paragraph on a familiar topic will be easier to process at all stages of comprehension".

Readability is connected to plain language. Plain language refers to how clear and concise texts are and how easy texts are to read and understand (Gilliver, 2015). Maass (2020: 12) writes that plain language was "a means to open expert contents for lay people, for example, by providing people without legal or medical training access to the respective expert communication". The idea behind readability is to match readers to appropriate texts, but also to ensure texts are understood by the wider public. DuBay (2004) writes that readability first became an important factor in education, and later in health and military services to combat errors due to difficult texts. For a text to be understood by the wider public, the level of difficulty should be the same reading level of 8th or 9th graders (Schriver, 2017).

Readability formulas

Readability can be assessed using a readability formula. According to DuBay (2004), there are over 200 readability formulas. Readability formulas estimate text difficulty based on mathematical equations and were created to predict the level of text difficulty without direct assessment from readers, which makes readability formulas efficient and objective (Mesmer, 2008). The most common measurements of readability are sentence length and word length. Klare (1974) states that other variables are not statistically significant for measuring the difficulty of a text. However, new developments in natural language processing and machine learning provide a more thorough analysis of language features and may improve readability tools (Crossley et al., 2019). Many readability formulas exist, such as the Lexile framework, Automated Readability Index (ARI), and FORCAST, but the most common formulas Flesch-Kincaid, Flesch Reading Ease (FRE), Dale-Chall, the Fog-Index, and SMOG (Simple Measure of Gobbledygook) will be focused on in the next section. New readability tools based on natural language processing will also be discussed. Lastly, a criticism of readability and readability formulas is presented.

Common readability formulas

The most popular readability formulas are the Flesch Reading Ease and Flesch-Kincaid formula, based on the formula developed by Rudolph Flesch in 1943 (Flesch, 1948). Other formulas such as Dale-Chall, Fog-index, and SMOG are widely used in readability estimations. These readability formulas assume word and sentence length determine the difficulty of a text. Although these two aspects are the basis of the formulas, the readability scores are calculated differently. The readability formulas result in a score and reading grade, based on age and education level. Table 3 explains how the readability formulas are calculated.

Formula	Calculation
Flesch-Kincaid	Grade = 0.39 (average words/sentence) +11.8 (average syllables/word) – 15.59
Flesh Reading ease	Grade = 206.835 – 1.015 (average words/sentence) – 84.6 (average syllables/word)
Dale-Chall	Grade = 0.1579 × (percent unfamiliar words) + 0.0496 × (word/sentence) + 3.6365
Fog-index	Grade = 0.4 [(average words/sentence) + 100 (percent of hard words)]
SMOG	Grade = 3 + (square root of the number of polysyllable word count)

***Table 3**. Readability formulas and their calculations*

The Flesch score falls between a number of 0-100, as shown in Table 4. The higher the number, the easier the text is to read. Academic texts are considered "Very difficult", and texts in this range score between 0-30. Low readability refers to difficult texts and high readability refers to easier texts. Hartley (2016) criticises the Flesh readability formulas for not accounting for the meaning of words, and readability programs based on the FRE count different things. Some counts syllables, and other counts the whole word. This provides different readability scores depending on which program is used (Hartley, 2016). Despite this, the Flesch score is widely used and accessible, partly due to its implementation in Microsoft word (Mesmer, 2008).

Flesch R.E. scores	Reading age	Difficulty	Example
90-100	10-11 years	Very easy	Comics
80-90	11-12 years	Easy	Pulp fiction
70-80	12-13 years	Fairly Easy	Popular novels
60-70	14-15 years	Average	Tabloid newspapers
50-60	16-17 years	Fairly difficult	Introductory textbooks
30-50	18-20 years	Difficulty	Undergraduate's essays
0-30	Graduate	Very difficult	Academic prose

Table 4. *Description of the Flesch Reading Ease (Hartley, 2016)*

Dale and Chall (1948) criticised the Flesch Reading Ease for being too difficult to use due to inconsistencies in the counting of words and syllables, which prompted the creation of the Dale-Chall formula. The main idea of the Dale-Chall formula is that a word list offers better predictions of readability than simple word length, since vocabulary is essential in readability assessment (Dale & Chall, 1948). A word list of the 3000 most common words known by 4^{th} graders is used to count the percentage of words not in the list. If the word does not appear on the list, it is a difficult word. The Dale-Chall formula results in a score between 4.9-9.9, with 9.9 being the most difficult (Table 5). The Dale-Chall formula has consistently been tested and developed, resulting in the New Dale-Chall formula with an updated word list, which was published in *Readability revisited: The new Dale-Chall readability formula* by Dale and Chall (1995), and is the most reliable and valid assessment tool of all classic readability formulas in predicting reading levels (Begeny & Greene, 2014; Mesmer, 2008).

Score	Notes
4.9 or under	easily understood by an average 4th-grade student or lower
5.0-5.9	easily understood by an average 5th or 6th-grade student

6.0-6.9	easily understood by an average 7th or 8th-grade student
7.0-7.9	easily understood by an average 9th or 10th-grade student
8.0-8.9	easily understood by an average 11th or 12th-grade student
9.0-9.9	easily understood by an average 13th to 15th-grade (college) student

Table 5. *Dale-Chall readability scores (Dale-Chall, 1948)*

Robert Gunning developed the Fog-Index in his book *The technique of clear writing* (1952). Gunning became interested in readability when he discovered many high school students were unable to read the school texts (DuBay, 2004). The Fog Index is calculated using sentence and word length and counting the percentage of hard words in the text that results in a reading grade.

McLaughlin (1974), dissatisfied with the Flesch Reading Ease and the Gunning Fog index, created SMOG (Simple Measure of Gobbledygook) and claimed it is simpler and more efficient. McLaughlin's (1974) main criticisms of the Flesch formula were the use of children's texts to assess readability and the lack of validity. McLaughlin (1974) conducted studies using British newspapers and magazines to assess adult British participants to decide the level of readability. The main idea is that polysyllabic words, words with three or more syllables, better assess text difficulty. The number of polysyllabic words determines a grade between 5-18. The steps to calculate the SMOG grade are explained in Table 6 below.

Smog Grading

1. Count 10 consecutive sentences near the beginning of the text to be assessed, 10 in the middle, and 10 near the end. Count as a sentence any string of words ending with a period, question mark or exclamation point.
2. In the 30 selected sentences count every word of three or more syllables. Any string of letters or numerals beginning and ending with a space or punctuation mark should be counted if you can distinguish at least three syllables when you read it aloud in context. If a polysyllabic word is repeated, count each repetition.
3. Estimate the square root of the number of polysyllabic words counted. This is done by taking the square root of the nearest perfect square. For example, if the count is 95, the nearest perfect square is 100,which yields a square root of 10. If the count lies roughly between two perfect squares, choose the lower number. For instance, if the count is 110, take the square root of 100 rather than that of 121
4. Add 3 to the approximate square root. This gives the SMOG Grade, which is the reading grade that a person must have reached if he is to understand fully the text assessed.

Table 6. *How to calculate the SMOG grade (McLaughlin, 1969).*

New developments in readability

The common readability formulas focus on word/syllable and sentence count, while some formulas use word lists to cover vocabulary. Crossley et al. (2019) criticise the traditional readability formulas since they are less predictive of text comprehension. Developments in natural language processing (NLP) better assess the difficulty of texts while accounting for multiple features of language (Benjamin, 2012; Crossley et al., 2019). Besides simple word and sentence counts, other aspects of texts are related to readability, comprehension, and reading speed, such as lexical sophistication (word difficulty), text cohesion (how the text is linked together), and syntactic complexity (grammatical difficulty) (Crossley et al., 2019). There are readability programs based on NLP available, but readability can be assessed by taking language features into account and a specific program or formulas is not always needed but they provide efficient tools for textual analysis.

Graesser et al. (2011) developed a readability program called the Coh-metrix, which measures 106 language features using NLP. Some of these features are descriptive such as number of paragraphs, words, or sentences. There are also several lexical diversity, cohesion, and syntactic complexity indices. Graesser et al. (2011) stress that text cohesion is an important aspect of readability and text comprehension, and the Coh-metrix measures textual aspects that are connected to how coherent texts are. Coh-metrix focuses on five categories of language such as narrativity, syntactic simplicity, word concreteness, referential cohesion, and deep cohesion. The program also provides the readability scores of FRE and Flesch-Kincaid Grade level, and the Coh-Metrix L2 Readability score. Graesser et al. (2011) state that a more in-depth formula provides guidelines on several aspects of the text since an aggregated score lacks indication of which language features contribute to text difficulty. They claim this approach to readability enables easier ways to modify texts:

> Scientific investigations of reading also benefit from automated measures of text characteristics. Automated measures allow researchers to sample and manipulate texts systematically to either target or control particular reading components in their investigations of comprehension. (Graesser et al., 2011: 223)

Crossley et al. (2011) compared the Coh-metrix to the Flesch Reading Ease and found the Coh-metrix performs better at predicting readability. The predictability of the Coh-metrix is also reaffirmed in Crossley et al. (2008). Coh-metrix is freely available online at cohmetrix.com as a web-based tool. Coh-metrix has been used to assess linguistic characteristics of students by Maamuujav et al. (2021).

Crossley et al. (2019) also developed the language assessing formulas CAREC and CARES, which use natural language processing methods to assess reading speed and text comprehension. Crossley et al. (2019) found that NLP can be the basis for new readability tools and programs since NLP-based tools perform better and have higher validity than the classic readability formulas. The CAREC and CARES programs are unavailable for public use since they are still in development to be more user-friendly, however the formulas can be used with existing NLP tools (Crossley et al., 2019).

Lu and Bluemel (2020) write that the future of linguistic analysis is Automated Language Assessment (ALA). "ALA systems identifies quantifiable features of language use that can both predict writing quality and use different dimensions of the writing construct" (Lu & Bluemel, 2020: 87). Two programs analysing lexical and syntactic complexity were developed by Lu (2010, 2012) called the Lexical Complexity Analyzer (LCA) and L2 Syntactic Complexity Analyzer (L2SCA) to assess second language learning. However, these programs have been used to analyse readability (Jin et al., 2021) and the language of university students (Ai & Lu, 2013; Lu & Ai, 2015). The LCA and L2SCA are freely available for download or as a web-based tool. Ai & Lu (2010) developed the web-based software since the downloadable version requires extensive knowledge of programming and UNIX systems. Vajjala and Meurers (2012) compared traditional readability formulas with Lu's lexical and syntactic tools and found that the LCA and L2SCA indices performed better in readability prediction. The LCA predicted 68.1% of reading levels accurately, whereas the L2SCA achieved an accuracy of 71.2% (Vajjala & Meurers, 2012). A detailed description of the LCA and L2SCA can be found in the Data collection and method chapter (pp. 48-50).

Criticism of readability formulas

Readability formulas were developed to save time and prevent errors when assigning reading grades. New technology has further increased speed and efficiency in readability assessments, and readability tools are freely available for download and online. The common readability formulas have been extensively used and researched (Mesmer, 2008), whereas new readability tools still need developing and testing and require more training and expertise in coding unless there is software readily available for ease of access. The readability formulas provide an objective assessment of text difficulty and are reliable (Mesmer, 2008). However, there are several criticisms regarding readability and the formulas.

The main criticism of readability formulas is validity and predictability. Begeny and Greene (2014) investigated the validity of the FRE, Dale-Chall, SMOG, and FOG formulas and found that they are not valid for assigning lower reading grades; however, they perform better when grading texts for more experienced readers. Predictability for these formulas is low, except for the Dale-Chall formula (Begeny & Greene, 2014). Janan and Wray (2014) write that readability involves both the text and the reader. However, many researchers have neglected the reader in their estimations. This may explain why readability formulas perform poorly in validity and reliability testing (Janan & Wray, 2014), which is a challenge new readability programs based on natural language processing face as well.

Bailin and Grafstein (2001) criticises the assumptions that longer words and longer sentences are more complex, which indicates that the texts are less readable. Bailin and Grafstein (2001) argue that this is not always the case. Some short words may be unknown to the reader, and shorter sentences may be difficult to decode. Nonetheless, the readability formulas would classify such texts as easy to read. Bailin and Grafstein (2001) argue that grammar, style, background knowledge, text coherence, and the interaction between these factors determine how difficult the text is to understand. None of the readability formulas can measure previous knowledge and whether the text is understood unless incorporating the reader.

Readability is not an exact science. Several variables must be considered to gauge the difficulty of texts. The most straightforward way to calculate readability, such as word and sentence count, is perhaps the most efficient. However, it fails to consider how several features of language create complexity. Despite assertions that word and sentence count is sufficient to evaluate readability, several studies show that the common readability formulas perform worse at predictability than new readability tools. The common readability formulas have faced extensive use and development, which is a curtesy that also need be extended to new developments of readability tools to establish them as valid forms of evaluation.

3. EARLIER RESEARCH

3.1 Readability and academic language

The academic paper is the most important channel to communicate research findings and theoretical advances. Academic language is generally difficult and academic texts score low on readability. Academic texts are described as precise, impersonal, and objective, using passive tense and complex structures (Hartley, 2008). Snow (2010) writes that features of academic language vary depending on discipline, topic, and mode. However, there are still core features that define most academic language, such as a high density of information bearing words, precision of expression, lexical sophistication (word difficulty), nominalisation (the transformation of a word into a noun), and complex grammar (Snow, 2010). Ventola (1996) states that scholars are given conflicting advice regarding academic writing. On the one hand, scientists are told to use short sentences and simple language; on the other hand, they are told to write complex and concise texts. Balance between simple and complex language is required, which can be difficult to achieve.

Academic texts are often the most difficult reading level, as exemplified earlier by the Flesch score and Dale-Chall formula. Studies on academic texts show that academic disciplines lie within a Flesch score of 0-30, as illustrated in Table 7 by Gazni (2011). Gazni (2011) also demonstrates the differences between disciplines, with research areas in medicine and chemistry being more difficult to read than mathematics, physics, engineering, and economics. However, all disciplines use difficult language and are very difficult to read when analysed by the Flesch Reading Ease.

Discipline	FRE score
Pharmacology and Toxicology	12.6
Multidisciplinary	13.4
Chemistry	13.8
Clinical Medicine	14.6
Microbiology	14.8
Neuroscience and Behavior	14.8
Biology and Biochemistry	14.8
Social Science, General	15.0
Environment/Ecology	15.4
Molecular Biology and Genetics	15.4
Psychiatry/Psychology	16.2
Immunology	16.6
Materials Science	16.8
Plant and Animal Science	17.2

Geosciences	18.2
Agricultural Sciences	18.6
Computer Science	19.2
Economics and Business	19.6
Engineering	19.8
Physics	22.6
Space Sciences	25.4
Mathematics	25.6
average	17.3

Table 7. *Average FRE scores of 22 disciplines, ranging from lowest to highest (Gazni, 2011)*

Studies on academic language and readability using new readability tools are few. Lu (2012), who developed language assessment tools for the evaluation of lexical and syntactic complexity, has mapped the lexical complexity of college-level English among non-native students. Table 8 summarises the average lexical complexity. Ai and Lu (2013) and Lu and Ai (2015) mapped the syntactic complexity of native English speakers' university essays, which can be found in Table 9. These studies set out to measure the complexity of language among non-native and native speakers to get an idea of the level of proficiency among students. These studies were not originally used to assess readability; however, these language indices provide an indication of the level of complexity in academic contexts that affect how readable the texts are. The analysis present for instance how many difficult words are used and the average sentence length, which is a feature the Flesch scores lack. More details on the lexical and syntactic complexity can be found in the Data collection and methods chapter (pp. 48-50).

Lexical complexity	Code	Mean
Lexical density	LD	0.41
Lexical sophistication 1	LS1	0.23
Lexical sophistication 2	LS2	0.26
Verb sophistication 1	VS1	0.07
Verb sophistication 2	VS2	0.31
Corrected verb sophistication	CVS1	0.33
Number of different words	NDW	119.83
NDW (first 50 words)	NDWZ	34.88
NDW (expected random 50)	NDWERZ	36.83
NDW (expected sequence 50)	NDWESZ	34.23
Type-token ratio	TTR	0.41
Mean segmental TTR (50)	MSTTR	0.69
Corrected TTR	CTTR	4.94
Root TTR	RTTR	6.99
Bilogarithmic TTR	LOGTTR	0.84

Uber index	UBER	26.15
Lexical word variation	LV	0.57
Verb variation 1	VV1	0.58
Squared VV1	SVV1	13.42
Corrected VV1	CVV1	2.56
Verb variation 2	VV2	0.19
Noun variation	NV	0.59
Adjective variation	ADJV	0.11
Adverb variation	ADVV	0.04
Modifier variation	MODV	0.15

Table 8. *Average lexical complexity of college-level English oral narratives (Lu, 2012)*

Syntactic complexity	Code	Ai & Lu (2013)	Lu & Ai (2015)
Mean length of sentence	MLS	19.15	19.60
Mean length of clause	MLT	17.07	17.31
Mean length clause	MLC	9.94	10.09
Sentence complexity ratio	C/S	-	1.97
Verb phrases per T-unit	VP/T	-	2.34
T-unit complexity ratio	C/T	-	1.73
Dependent clause ratio	DC/C	0.40	0.40
Dependent clauses per T-unit	DC/T	0.73	0.73
Sentence coordination ratio	T/S	1.21	1.13
Complex T-unit ratio	CT/T	-	0.51
Coordinate phrases per T-unit	CP/T	0.43	0.43
Coordinate phrases per clause	CP/C	0.25	0.25
Complex nominals per T-unit	CN/T	2.09	2.09
Complex nominals per clause	CN/C	1.22	1.22

Table 9. *Average syntactic complexity of college-level essays*

There are concerns that research articles are too difficult to read. If the text is written to be accessible to the general audience, the Flesch scores should be between 60-70, but academic language scores are far below that as demonstrated by Gazni (2011). Worrall et al. (2020) stress the importance of information being readable to the wider public when misinformation cause adverse effects, especially during times such as the Covid-19 pandemic, and when people increasingly rely on having their informational needs met online. Readability is a matter of accessibility, but not everyone agrees that increasing the readability of academic information or health information solves the problem. Hosseini and Akbarzadeh (2021) express concerns that simplifying academic language may further cause misunderstanding and misinformation. Basch et al. (2021) write that increasing the level of readability is important to ensure that scientific discoveries are accessible to everyone, regardless of educational level and is a democratic principle, as initially concluded in Basch et al. (2020).

However, research articles have become less readable over the years, decreasing accessibility (Plavén-Sigray et al., 2017). Plavén-Sigray et al. (2017) offer three explanations as to why. First, the number of authors per article has increased. Secondly, the decrease in readability can be explained by the increased use of scientific jargon, which contributes to specialised language used in scientific texts that is not commonly known or used by a layperson. Lastly, the amount of research published, and the expanding and specialised scientific knowledge require complex language to express precise and specific information (Plavén-Sigray et al., 2017). Efforts to have been made to combat the low readability of research articles, especially in medical sciences. According to Smith et al. (2017), medical abstracts use too difficult language, resulting in information regarding health and medicine being inaccessible to patients, which is why plain language summaries are used in medical sciences to combat difficult language. Stricker et al. (2020) investigate the readability of psychology abstracts and their respective plain language summaries. They found that plain language summaries are easier to read than scientific abstracts, fulfilling their purpose.

Issues with low readability occur in academic contexts as well. King (1976) found that academic language is often too difficult for undergraduates to comprehend, and these concerns are still present. Snow (2010) writes that the complex structure and vocabulary of academic texts may “disrupt reading comprehension and block learning “ among students. Brown et al. (2019) discovered that complex language inhibits cognitive abilities, which indicates that it takes longer for students to process tasks when exposed to complex scientific language. Key features of academic language such as nominalisation pose problems for students and their understanding of scientific texts (Hao & Humphrey, 2019). Furthermore, Plavén-Sigray et al. (2017) state that replicability of results is affected due to misunderstandings when reading methods chapters of articles that score low on readability. Warren et al. (2021) reiterate similar concerns and write that low readability can limit the impact of research. They argue that abstraction, passive writing, and technical language negatively affect the understanding of research, while researchers simultaneously want to reach a wider audience. However, the notion that academic language must be overly complex lingers since it contributes to a good impression of the writer, and researchers forget that readers may have less knowledge about scientific topics, called the "curse of knowledge" which contributes to lack of clarity and difficult texts (Warren et al., 2021).

The academic abstract

The abstract is the first piece of text that is read, after the title of the paper (Hartley, 2008) and has limited space, which stresses the importance of conveying relevant information efficiently. Each institution, journal, or discipline have their guidelines for writing abstracts. There are two abstract types: structured and traditional. The structured abstract was introduced in the 1980s in the medical sciences and is divided into subheadings such as background, method, results, and conclusion (Hartley, 2004). The traditional abstract lacks such divisions and is a running text with generally one paragraph. The content of the abstract is the same regardless of type and length. The abstract must give an overview of the aim of the article, methods used, general results, and main conclusions. The main idea is to provide the most necessary information for the reader to decide if the article is relevant to their needs (Weil, 1970).

Abstracts are generally more difficult to read than the full text of the research articles (Yeung et al., 2018). Lei and Yan (2016) also discovered that abstracts are less readable than the rest of the article. This may be due to the limited space to express the main ideas and findings of the article, resulting in lengthy and complex sentences where a lot of information is communicated (Lei & Yan, 2016). Cohen et al. (2010) also found that the structure of abstracts is different from the full body text of the article. However, abstracts have shorter sentences and use fewer passive structures than the full text of the article, which indicates that abstracts are less complex than the rest of the article text (Cohen et al., 2010).

3.2 Readability and scientific impact

Many studies on readability and scientific impact use the Flesch Reading Ease to estimate the difficulty of academic texts (Gazni, 2011; Yeung et al., 2018; Dolnicar & Chapple, 2015). Some research also employ multiple readability formulas (Lei & Yan, 2016; Didegah & Thelwall, 2013; Sawyer et al., 2008). Previous research exhibit inconclusive results whether difficult texts influence scientific impact. Disciplinary differences exist and some researchers suggest that articles should be easier to read to increase accuracy in reporting research findings and increase accessibility (Yeung et al., 2018; Dolnicar & Chapple, 2015).

As mentioned, there are differences between disciplines and how readability affects citation counts. Gazni (2011) investigates the readability of abstracts from the top five institutions in the world, focusing on 22 disciplines. Gazni (2011) found that abstracts are very difficult to read, and certain disciplines are more difficult to read than others as shown in Table 7. Gazni (2011) also found that articles with less readable abstracts were cited more. This can be explained by the previous knowledge of the reader. Scientific texts with highly technical language poses few problems for experts in the same field. Conversely, Didegah and Thelwall (2013) found that the readability of abstracts affected citations negatively in biology and biochemistry, which indicates difficult texts in these disciplines receive fewer citations. Furthermore, Lei and Yan (2016) investigated the readability of research articles in library and information science and found no correlation between readability scores (Flesch Reading Ease and SMOG) and citations between 2003-2012. Sienkiewicz and Altmann (2016) studied lexical and sentiment factors affecting scientific impact of research articles in multiple disciplines and journals. They found that longer abstracts positively affects how many citations articles receive. Additional findings were that the textual complexity of abstracts correlates with citations. However, the effects were small and insignificant, and the conclusion is that textual features have no overall effect on scientific impact (Sienkiewicz & Altmann, 2016).

Other studies have focused on the top cited articles in specific disciplines. Yeung et al. (2018) investigate the 100 most cited neurosciences and neuroimaging papers and their readability. They found that readability is low, with an average Flesch Reading Ease score of 15.70, which was slightly higher than Gazni's (2011) results (14.8). However, the implication is that difficult language leads to more citations. Yeung et al. (2018) conclude that research articles in this field should be easier to read since there is evidence to suggest that the accuracy of media reporting of high impact articles is low. Dolnicar and Chapple (2015) support that research articles should aim for higher readability. They found that highly cited tourism articles are harder to read, and readability has become lower over time. Dolnicar & Chapple (2015) suggest that a Flesch score of 40 should be the aim of research articles. However, increasing readability is difficult and time consuming (Klare, 1974; Dolnicar & Chapple, 2015), and if less readable texts receive more citations it fails to incentivise scholars to write more readable texts.

Other readability formulas have also been used to analyse readability. McCannon (2019) studies the relationship between citations and readability of articles published in American Economic Review using the Linsear Write metric, developed by the U.S. Air Force and it mainly focuses on the number of syllables per words used. The results suggest that less readable articles receive fewer citations (McCannon, 2019). Sawyer et al. (2008) found that award-winning articles in marketing are more readable based on analysis of Flesch scores, SMOG, Fog-Index, ARI and Lix index. This indicates that economic disciplines favour texts with high readability. Dziubaniuk et al. (2021) add that articles using complex language in business marketing research are left unread by managers, which indicates that research with too complex language is potentially less influential.

3.3 Readability and online attention

Few studies have been conducted on readability and online attention. Results indicate that the difficulty of texts and linguistic features influence online attention to various degrees. However, these studies focus on a few journals or individual research areas. The readability analyses are based on common readability formulas and new tools and language assessment methods as well.

Jin et al. (2021) studied the relationship between online attention and readability of abstracts using the Lexical Complexity Analyzer (Lu, 2012) and L2 Syntactic Complexity Analyzer (Lu, 2010). They discovered that certain lexical and syntactic indices influence the level of online attention. Lexical sophistication had a negative effect on Altmetric Attention Scores, which Jin et al. (2021) theorise could be due to non-scholars opting to read articles using less difficult words. However, no evidence of non-scholars sharing these articles are presented. The other lexical index influencing online attention was lexical variation. Higher word variation increased online attention. Regarding syntactic complexity indices, Jin et al. (2021) found that verbal phrases had a negative effect on online attention, whereas complex nominals had a positive impact. The results show that different features of language influence online attention in different ways. Ruano et al. (2018) also found that readability affects social media attention. Articles with lower readability were less visible on social media (Ruano et al., 2018). However, they only studied a specific research area (psoriasis treatment) and results cannot be generalised across more disciplines and journals. Same conclusions can be drawn about the research conducted by Jin et al. (2021) since they only focused on the journal *Science*.

Research on readability and online attention also focus on usage metrics related to downloads, bookmarks, and highly browsed articles. The conclusions are that there are differences between highly bookmarked articles and highly downloaded articles. Chen et al. (2020) analysed linguistic characteristics of articles published in PLoS journals such as title length, abstract length, sentence length, lexical diversity, lexical density, and lexical sophistication of highly browsed and downloaded articles. They discovered that title length and sentence length affected the number of downloads and that the abstracts were longer in highly downloaded articles, but not enough to be statistically significant and there were differences between the different PloS journals (Chen et al., 2020). Guerini et al. (2012), using the Fog-index and Flesch scores, discovered differences between the most bookmarked and the most downloaded articles. The most bookmarked articles had more difficult abstracts, whereas the most downloaded articles were more readable. Guerini et al. (2012) explain that articles that require less initial understanding will be downloaded immediately, whereas articles with lower readability require more understanding and will be bookmarked for future reading. However, cited papers are less connected to readability, which leads Guerini et al. (2012) to conclude that it is ultimately the style and content of the article that matters, not text difficulty. The results from earlier studies investigating readability and online attention are thus inconclusive and more research is needed to fully understand the implications readability may have for online attention.

4. DATA COLLECTION AND METHOD

The aim of this thesis was to investigate the readability of abstracts and assess the relationship between abstract readability and the online attention of research articles published in 2018 affiliated with Finnish Universities. The objective was to map the readability of the research article abstracts across disciplines using readability tools based on natural language processing and analyse the correlation between lexical and syntactic complexity indices of abstracts and the Altmetric Attention Score.

The research questions are:

1. What is the lexical and syntactic complexity of research article abstracts? Are there differences between research areas?
2. What is the relationship between readability and online attention? Are there differences between research areas?

Abstracts were collected from Web of Science and Altmetric Attention Scores were collected from Altmetric Explorer by provided by Altmetric.com. Readability was analysed with the web-based versions of the Lexical Complexity Analyzer (Lu, 2012) and the L2 Syntactic Complexity Analyzer (Lu, 2010). The study used quantitative methods since research in altmetrics often employ quantitative methods to collect and analyse data from relevant databases and sources of altmetrics (Holmberg, 2015). The relationship between readability variables and Altmetric Attention Scores was analysed using Spearman's correlation coefficient (r_s).

Descriptive statistics were used to analyse and describe the results from the LCA and L2SCA analysis, and to describe altmetric data. The mean (M) values were included, as was the standard deviation (SD) to explain the spread and dispersion of data. Median values and maximum values were also presented when describing the altmetric data, to further explain the spread and skewness of data. Inferential statistics were used to test the association between variables (Bhattacherjee, 2012). The statistical significance of the correlation analysis was set to a significance level (α) = 0.05, which is most common in social sciences (Bhattacherjee, 2012). Material, data collection, and methods are explained further in the following sections.

Research article abstracts

Research articles affiliated with Finnish universities published in 2018 were collected for analysis. Metadata and abstracts from 9 482 articles were collected in January 2022 from Web of Science. To limit the data, the top 15 research areas, as categorised by Web of Science, with the most published articles were selected. 15 research areas were also selected to include a variety of disciplines. Only articles written in English were selected since the readability software is developed for texts written in English. Furthermore, only the top 15 research areas with the most published articles were selected for analysis. Web of science tracks 151 research areas. Articles with missing DOI and missing abstracts were removed. Abstracts with less than 50 words were also removed due to the readability software only accepting texts with a minimum of 50 words. DOI is used to search for articles in Altmetric Explorer, which is why articles with missing DOI were removed. The total data consists of 9 272 abstracts. The research areas and the number of abstracts per research area are presented in Table 10.

Research area	Code	N	Removed	Words		Sentences	
				mean	SD	mean	SD
Astrophysics and astronomy	AA	380	0	234.44	89.74	9.57	4.61
Business economics	BE	623	14	208.01	57.07	8.66	2.43
Biochemistry and biology	BIO	390	1	171.93	61.90	7.28	2.74
Chemistry	CHEM	858	19	182.07	64.97	7.39	2.67
Computer science	CS	451	28	192.31	66.64	8.47	2.94
Ecology	ECO	1002	16	235.69	69.38	9.73	3.14
Education	EDU	321	21	178.12	58.87	7.90	2.90
Engineering	ENG	1018	18	195.92	61.04	8.66	2.88
Environmental and occupational health	EOH	267	6	242.42	66.64	10.54	3.12
Mathematics	MATH	292	48	179.41	56.39	7.66	2.75
Materials science	MS	721	19	135.11	62.09	5.71	2.50
Neuroscience and neurology	NN	461	5	242.08	63.12	10.83	3.50
Physics	PHY	1085	7	172.87	65.29	6.88	2.53
Psychology	PSY	373	4	196.41	55.49	8.49	2.77
Science technology	ST	1030	4	197.24	60.78	8.44	2.83
All		9272	210	204.09	70.66	8.53	3.25

Table 10. *Number of abstracts and removed data, and average number of words and sentences per research area*

Altmetric Attention Score

Articles with at least 1 Altmetric attention score were selected for analysis. Articles with only Mendeley readership count were removed since it is not counted towards the Altmetric Attention Score. The Altmetric Attention Score is based on three factors: volume, sources, and authors. These factors are calculated using a weighted count based on the assumption that different altmetric indicators bring different levels of attention, (see chapter 2.3 on page 21), resulting in what is called the altmetric donut, visualised in Figure 2. For instance, one news article weighs the most with a score of 8, blog posts have a score of 5, a Wikipedia entry has a score of 3. One tweet result in 0.25 points. The Altmetric Attention Score must be an even number, which means the score of an article with one tweet has an attention score of 1 and an article with three tweets also has a score of 1.

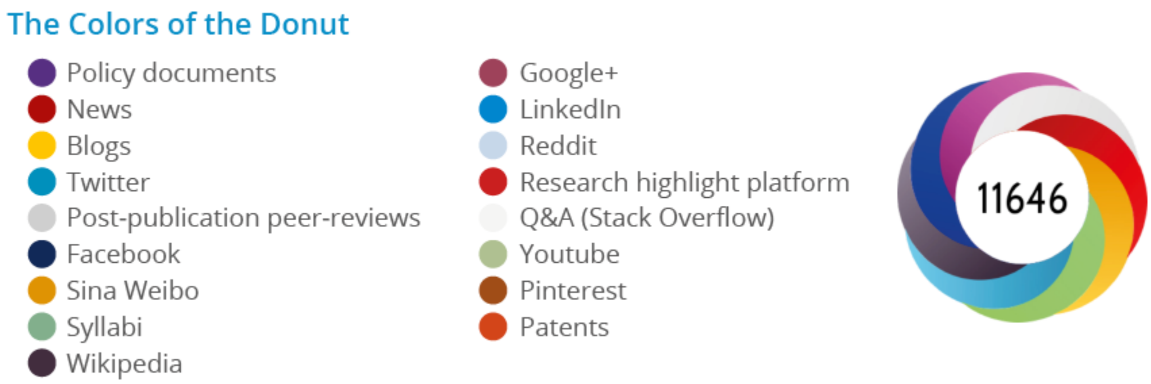

Figure 2. *The altmetric sources that comprise the altmetric donut (Altmetric.com)*

Altmetric Attention Scores (AAS) were collected in January 2022 from Altmetric explorer provided by Altmetric.com. The total amount of articles collected was 5 604. Table 11 shows the number of articles with an altmetric event and the average AAS. The altmetric data was heavily skewed, with most articles having low AAS and few articles receiving the most online attention. 60% of articles had at least one AAS. Astronomy astrophysics (89%) had the highest percentage and Engineering (30%) had the lowest. Science technology and other (M = 40.32, SD = 132.72) had the highest average AAS, whereas Mathematics (M= 4.41, SD = 10.82) had on average the lowest AAS. The descriptive summary in Table 12 shows that Twitter, blogs, news, and Wikipedia had the highest coverage of all altmetric sources. Out of these, Twitter (M = 13.25, SD = 65.12) had the highest coverage.

Research area	Mean	SD	Median	maxAAS	NAAS	N	NAAS/N %
AA	12.28	40.07	5	637	337	380	89 %
BE	7.77	17.42	3	252	288	623	46 %
BIO	22.17	108.92	3	1523	301	390	77 %
CHEM	6.30	17.22	2	215	460	858	54 %
CS	7.79	15.31	3	111	176	451	39 %
ECO	20.29	60.27	5	1006	731	1002	73 %
EDU	9.94	35.35	3	415	171	321	53 %
ENG	9.08	36.17	2	396	301	1018	30 %
EOH	15.38	38.15	4	455	217	267	81 %
MATH	4.41	10.82	1	76	92	292	32 %
MS	9.18	32.12	2	396	313	721	43 %
NN	24.19	78.15	4	947	392	461	85 %
PHY	7.72	26.72	1	459	773	1085	71 %
PSY	15.01	32.16	5	273	274	373	73 %
ST	40.32	132.72	4	1483	778	1030	76 %
Total					5604	9272	60 %

Table 11. *Descriptive summary of Altmetric attention scores and number of articles with an altmetric event. maxAAS = the highest altmetric attention score of the data, NAAS = number of articles with an altmetric attention score, N = Number of article abstracts collected in total, NAAS/N % = percentage of articles with an altmetric attention score*

Altmetric sources	Mean	SD
Altmetric Attention Score	16.79	67.92
News	1.10	6.27
Blog	0.28	1.25
Policy	0.06	0.79
Patent	0.04	0.29
Twitter	13.25	65.12
Facebook	0.37	1.33
Wikipedia	0.10	2.76
Google+	0.08	0.97
Reddit	0.03	0.25
F1000	0.01	0.15
Video	0.02	0.15

Table 12*. Descriptive summary of altmetric sources*

Readability analysis

Traditional readability formulas calculate an aggregated number based on the average number of words and sentences in a text. New developments in readability call for more aspects of language to be analysed to better understand readability (Graesser et al., 2011; Crossley et al., 2019). Natural language processing enables efficient analysis of multiple linguistic aspects of texts. There are measurements of language difficulty that take more detailed lexical and syntactical features into account. Lu's (2012) Lexical Complexity Analyzer (LCA) provides 25 different lexical indices, and the L2 Syntactic Complexity Analyzer (L2SCA) analyses 14 indices of syntactic complexity (Lu, 2010). These programs have been used to assess the readability of university-level texts (Jin et al., 2021; Ai & Lu, 2015). Lexical and syntactic complexity perform better at predicting readability than traditional readability formulas (Vajjala & Meurers, 2012)

The LCA and L2SCA are freely available, as downloadable programs for UNIX systems, or as web-based software tools. The web-based versions were used in this thesis, since they streamline the natural language processes and remove the need to tag, lemmatise and program the code yourself (Ai & Lu, 2010), which saves time and reduces errors. The web-based software uses the Stanford POS (part of speech) and the Morpha lemmatizer (Ai & Lu, 2010). A part of speech tagger identifies words such as nouns, verbs, and adjectives (Stanford NLP Group, n.d.). Lemmatization removes affixes and gives the base form of the word and is explained by Manning et al. (2008) as the following:

> Lemmatization usually refers to doing things properly with the use of a vocabulary and morphological analysis of words, normally aiming to remove inflectional endings only and to return the base or dictionary form of a word, which is known as the *lemma.* (Manning et al., 2008).

Lexical Complexity Analyser

Lu (2012) developed a computational system to automate the analysis of lexical richness in three dimensions called the Lexical Complexity Analyser (LCA). Lexical refers to words and vocabulary. The three dimensions of lexical richness are lexical density, lexical sophistication, and lexical variation. Lexical density refers to the total number of lexical words to the number of words in a text. Lexical sophistication refers to lexical rareness or the number of unusual or advanced words in a text. Lu (2012) counts lexical sophistication by counting how many words are not found in the top 2000 most common words in the British National Corpus. Lexical variation, also called lexical diversity, refers to the range of vocabulary found in a text (Lu, 2012). The LCA measures 19 different metrics of lexical variety. In total, the LCA has 25 lexical indices, as described in Figure 3.

Number	Measure	Code	Formula
Lexical density			
1	Lexical density	LD	N_{lex}/N
Lexical sophistication			
2	Lexical sophistication-I	LS1	N_{slex}/N_{lex}
3	Lexical sophistication-II	LS2	T_s/T
4	Verb sophistication-I	VS1	T_{sverb}/N_{verb}
5	Corrected VS1	CVS1	$T_{sverb}/\sqrt{2N_{verb}}$
6	Verb sophistication-II	VS2	T^2_{sverb}/N_{verb}
Lexical variation			
7	Number of different words	NDW	T
8	NDW (first 50 words)	NDW-50	T in the first 50 words of sample
9	NDW (expected random 50)	NDW-ER50	Mean T of 10 random 50-word samples
10	NDW (expected sequence 50)	NDW-ES50	Mean T of 10 random 50-word sequences
11	Type-token ratio	TTR	T/N
12	Mean Segmental TTR (50)	MSTTR-50	Mean TTR of all 50-word segments
13	Corrected TTR	CTTR	$T/\sqrt{2N}$
14	Root TTR	RTTR	$T/\sqrt{N}$
15	Bilogarithmic TTR	LogTTR	$LogT/LogN$
16	Uber Index	Uber	$Log^2N/Log(N/T)$
17	Lexical word variation	LV	T_{lex}/N_{lex}
18	Verb variation-I	VV1	T_{verb}/N_{verb}
19	Squared VV1	SVV1	T^2_{verb}/N_{verb}
20	Corrected VV1	CVV1	$T_{verb}/\sqrt{2N_{verb}}$
21	Verb variation-II	VV2	T_{verb}/N_{lex}
22	Noun variation	NV	T_{noun}/N_{lex}
23	Adjective variation	AdjV	T_{adj}/N_{lex}
24	Adverb variation	AdvV	T_{adv}/N_{lex}
25	Modifier variation	ModV	$T_{adj} + T_{adv}/N_{lex}$

T = Number of word types; N = Number of word tokens. The subscript under T or N denotes the number of types of tokens of particular category of words: lex = lexical words; slex = sophisticated lexical words; s = sophisticated words; sverb = sophisticated verbs; adj = adjectives; adv = adverbs

Figure 3. *Description of the Lexical Complexity Analyser indices from Jin et al. (2021).*

L2 Syntactic Complexity Analyser

The L2 Syntactic Complexity Analyser (L2SCA), also developed by Lu (2010), measures 14 indices of syntactic complexity. Syntax refers to grammar and sentence structure. The syntactic complexity indices are divided into four subdivisions: length of production unit, sentence complexity, subordination, coordination, and particular structures. Length of production unit refers to sentence, clause and T-Unit length. A T-unit can be a sentence but is defined as the shortest grammatical part that a sentence can be split into. Subordination measures the number of dependent clauses in a text and the number of clauses per T-unit. Coordination refers to the number of phrases with linking words. Particular structures refer to how much nominalisation is used, and the number of verb phrases used. The 14 indices and their calculations are explained in Figure 4.

Measure	Code	Definition
Type 1: Length of production unit		
Mean length of clause	MLC	# of words / # of clauses
Mean length of sentence	MLS	# of words / # of sentences
Mean length of T-unit	MLT	# of words / # of T-units
Type 2: Sentence complexity		
Sentence complexity ratio	C/S	# of clauses / # of sentences
Type 3: Subordination		
T-unit complexity ratio	C/T	# of clauses / # of T-units
Complex T-unit ratio	CT/T	# of complex T-units / # of T-units
Dependent clause ratio	DC/C	# of dependent clauses / # of clauses
Dependent clauses per T-unit	DC/T	# of dependent clauses / # of T-units
Type 4: Coordination		
Coordinate phrases per clause	CP/C	# of coordinate phrases / # of clauses
Coordinate phrases per T-unit	CP/T	# of coordinate phrases / # of T-units
Sentence coordination ratio	T/S	# of T-units / # of sentences
Type 5: Particular structures		
Complex nominals per clause	CN/C	# of complex nominals / # of clauses
Complex nominals per T-unit	CN/T	# of complex nominals / # of T-units
Verb phrases per T-unit	VP/T	# of verb phrases / # of T-units

Figure 4. *The 14 indices of syntactic complexity from Jin et al. (2021).*

Correlation analysis

The Spearman correlation coefficient (r_s) was used to assess the relationship between complexity indices and Altmetric Attention Score. This method ranks data based on “the ranks of the observation and not on the numerical values of the data” (Kothari, 2004: 302). If two data points have the same value, the average of the ranks will be calculated.

Altmetric data are often highly skewed. Therefore, a ranked correlation method such as Spearman is better, since Pearson correlation is sensitive to outliers (Thelwall, 2020). The Spearman (r_s) correlation coefficient formula:

$$r_s = 1 - \left\{\frac{6\sum D^2}{n(n^2-1)}\right\}$$

Where *n* is the number of paired observations, *D* is the difference between these paired observations, and the differences are squared to obtain the total of differences $\sum D^2$ (Kothari, 2004). Spearman r_s can be calculated using the RANK function in Microsoft Excel and then using Pearson's correlation coefficient on the ranked data (Statology, 2020), which was the method used for this thesis. Spearman r_s results in a number between 0 (no relationship), 1 (perfect positive relationship), or -1 (perfect negative relationship).

5. RESULTS

5.1 Readability analysis

Lexical complexity

Lexical density. Lexical density measures the number of lexical items of a text. The average number of lexical items per number of words for all research areas was M = 0.59, SD = 0.05. Overall, the spread of lexical density was small. Biochemistry and biology and computer science scored slightly above average and had the highest lexical density. Astrophysics and astronomy and mathematics scored lowest in lexical density. The level of Lexical density across research areas is visualised in Figure 5. All Lexical density data can be found in Appendix 1.

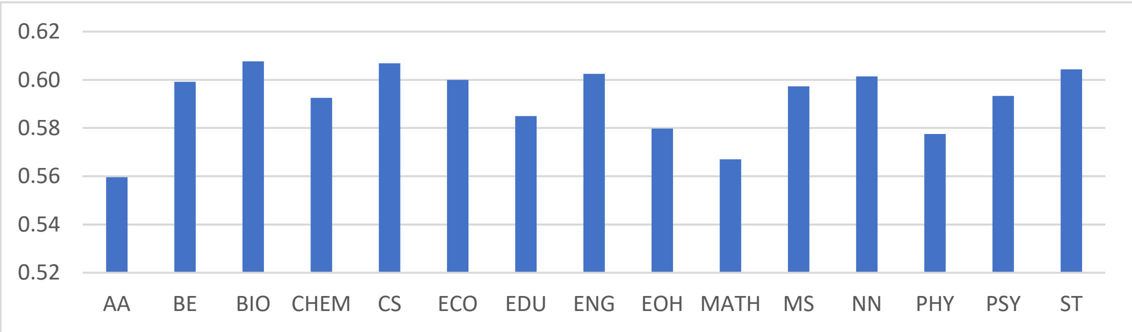

Figure 5. *The Lexical density of research areas*

Lexical sophistication. Lexical sophistication and verb sophistication, which refers to word difficulty, showed higher dispersion across research areas than lexical density, however, the variance was still small. The research areas with the highest number of difficult words (LS1 and LS2) were biochemistry and biology, chemistry, physics, materials science, neuroscience and neurology, and astrophysics and astronomy. The research areas with the lowest number of difficult words were business economics, education, psychology, computer science, and environmental and occupational health.

The indices with the higher dispersion of data in verb sophistication were VS2 and CVS1. The research areas with the highest number of difficult verbs (VS2, CVS1) were biochemistry and biology, science technology, astrophysics and astronomy, ecology, and neurosciences and neurology. Research areas that scored lower in word difficulty also scored low in verb difficulty such as psychology, environmental and occupational health, business economics, education, and computer science. Mathematics scored the lowest of all research areas in verb difficulty as shown in Figure 6. All data can be found in Appendix 1.

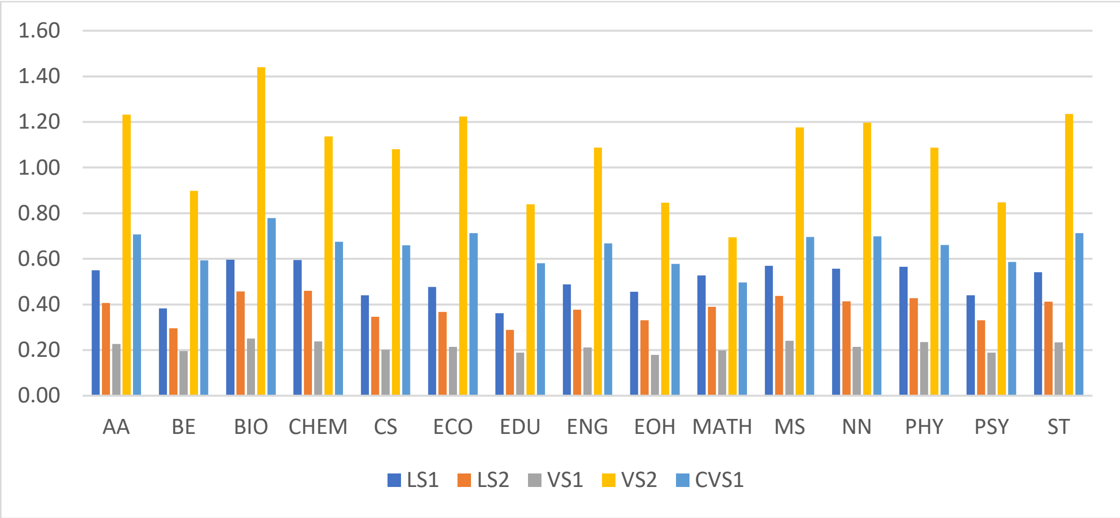

Figure 6. *Lexical sophistication indices across research areas*

Lexical variation. The lexical variation indices with the highest dispersion and variance were the number of different words indices (NDW, NDWZ, NDWERZ, and NDWESZ, which corresponds to NDW, NDW-50, NDW-ER50, AND NDW-ES50 in table 3), UBER index, and one verb variation index (VV1). Noun, adjective, adverb, and modifier variation was consistent across all research areas, which is summarised in Table 5 in the appendices. Chemistry, biochemistry and biology, science technology, and materials science scored the highest on the UBER index. In contrast, environmental and occupational health, psychology, neuroscience and neurology, mathematics, and education scored low. The research areas with the highest verb variation (VV1) were ecology, computer science, neuroscience and neurology. Mathematics scored the lowest again, with chemistry, physics, and materials science following. The number of different words, Type-token ratio and UBER indices, lexical word variation and verb variation can be found in Appendices 2-4.

The average number of different words was M = 110.63 SD = 29.76. Neuroscience and neurology, ecology, Public environmental occupational health, and astrophysics and astronomy had the most diverse number of words (NDW) with over 120 different words per abstract. Mathematics was below the average, with the fewest number of different words (M = 78.30 SD = 26.09). Business economics, education, and physics scored on average under 100 different words per abstract. The other number of different word indices (NDWZ, NDWERZ, and NDWESZ) showed similar results, but science technology scored the highest in NDWZ. The index NDW is summarised in Figure 7, and indices NDWZ, NDWERZ, and NDWESZ are summarised in Figure 8.

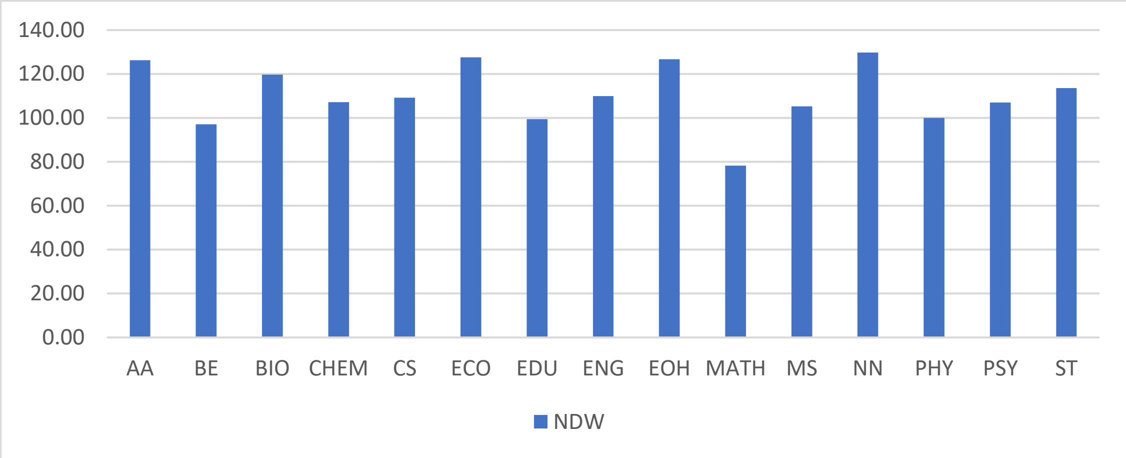

Figure 7. *Number of different words (NDW) across research areas*

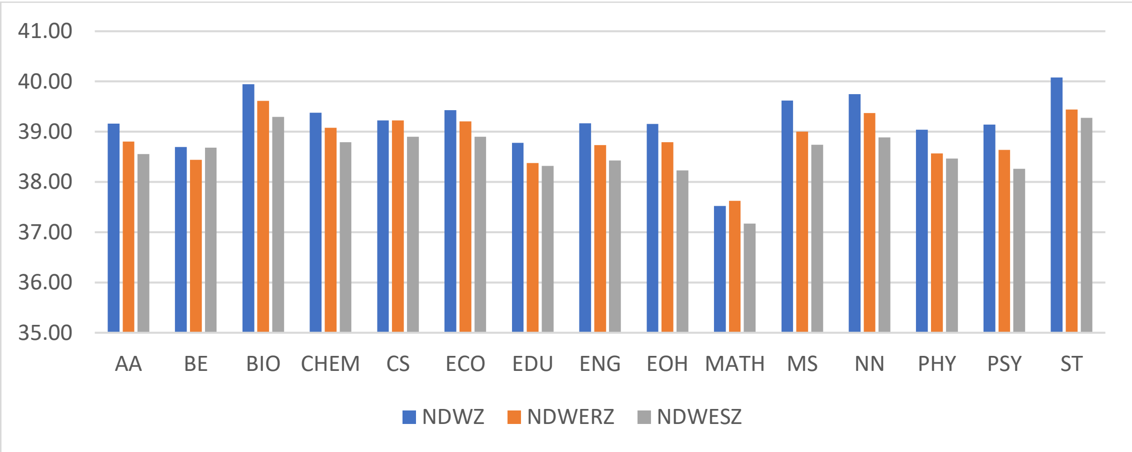

Figure 8. *Number of different words indices (NDWZ, NDWERZ, NDWESZ) across research areas*

Syntactic complexity

Length of production unit. Length of production unit covers the average number of words per sentence/T-unit/clause. The average sentence length was M = 24.67, SD = 8.41, the average T-unit was M = 23.72, SD = 8.52, and clauses were on average M = 17.56, SD = 7.12. The spread and variance of data were moderate. Mathematics had on average longer sentences and a high spread between data across all three indices. Astrophysics and astronomy, physics, and chemistry also had the longest sentences.

Engineering had on average the shortest sentences and T-units as visualised in Figure 9. other research areas with the shortest sentences were computer science, education educational research and engineering. Astrophysics and astronomy, mathematics, physics had the longest T-units and clauses. All data on the length of production unit can be found in Appendix 6.

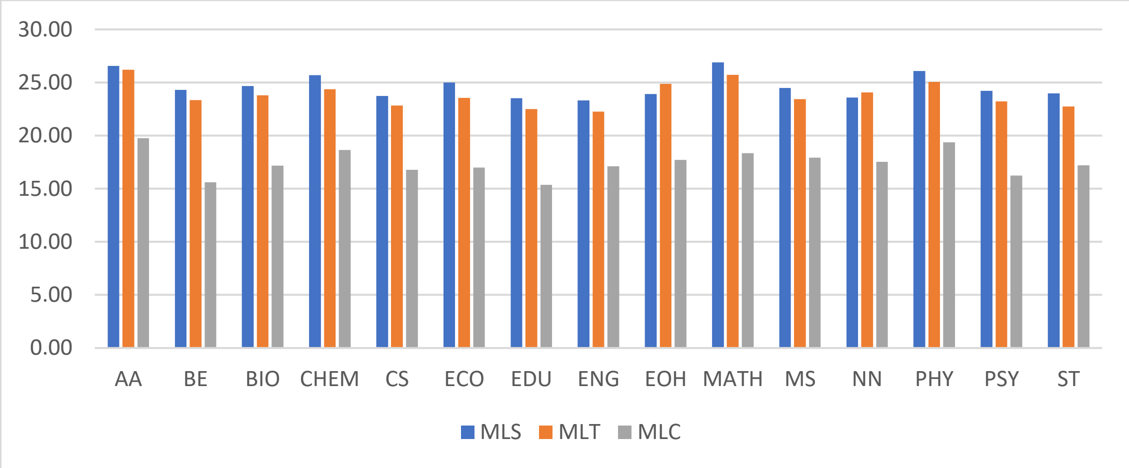

Figure 9*. Length of production unit indices across research areas*

Complex structures. The overall number of clauses per sentence (C/S) was M = 1.46, SD = 0.35, and had a small spread and variance across research areas. However, as shown in Figure 10, Business economics had on average the highest C/S. Education, mathematics, and psychology also scored high on sentence complexity. Neuroscience and neurology, engineering, astrophysics and astronomy, physics, materials science, and environmental and occupational health had the least complex structures. All data on complex structures can be found in Appendix 7.

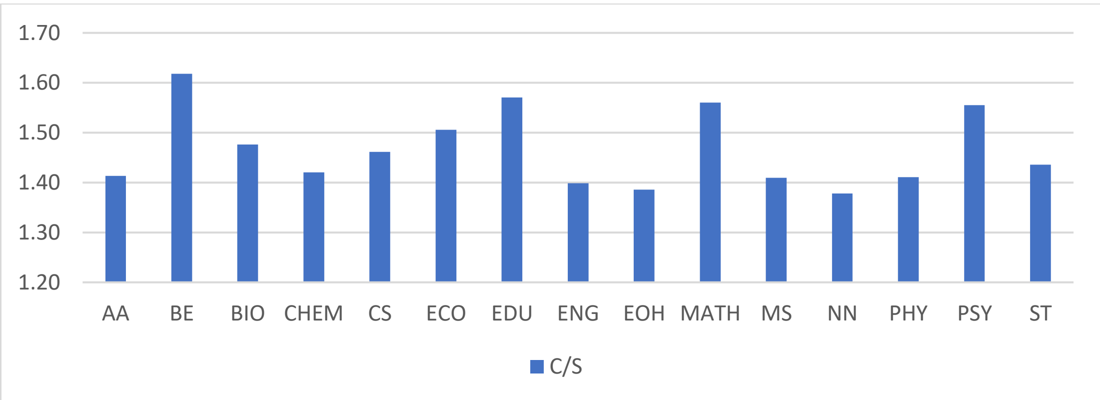

Figure 10. *Complex structures across research areas*

Subordination. Subordination indices, which refers to dependent clauses and T-unit complexity, had low spread and variance. Business economics had on average the highest t-units per clause (C/T), T-unit complexity ratio (CT/T) and dependent clauses per clause (DC/C) and T-unit (DC/T). Other research areas with high complexity in these indices were education, mathematics, psychology. Chemistry, engineering, science technology, and environmental and occupational health scored low in subordination as shown in Figure 11. All data on subordination can be found in Appendix 8.

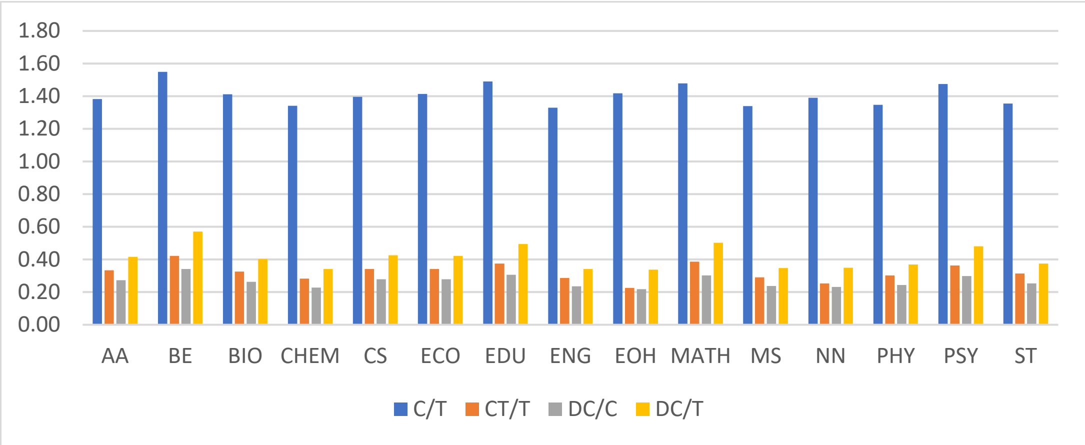

Figure 11. *Subordination indices across research areas*

Coordination. Coordinate phrases per clauses (CP/C) was on average M = 0.55, SD = 0.33 as shown in Figure 12. Environmental and occupational health had on average the most coordinate phrases. Coordinate phrases per T-units were M = 0.74, SD = 0.40, of which environmental and occupational health (M = 0.93, SD = 0.47) again had the most CP/T. Mathematics had on average the lowest CP/C and CP/T, whereas environmental and occupational health had the lowest sentence coordination ratio. Overall coordination was similar across research areas with small dispersion of data. All coordination indices and data points are summarised in Appendix 9.

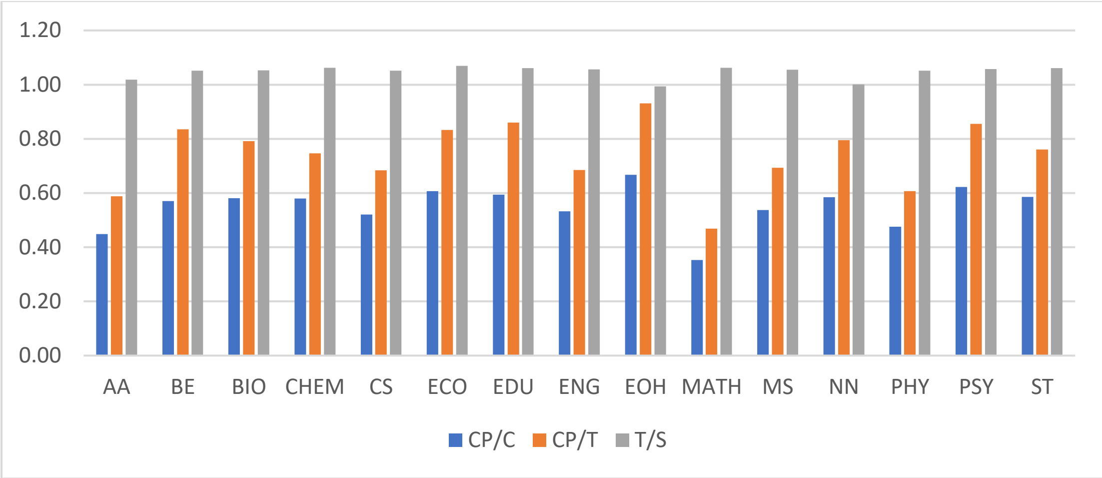

Figure 12. *Coordination indices across research areas*

Particular structures. Particular structures refers to complex nominals per clause (CN/C), complex nominals per T-unit (CN/T), and verb phrases per T-unit (VP/T). As visualised in Figure 13, the variance of these indices between research areas was small. Most research areas were above 2.50 average complex nominals per clause (CN/C), expect computer science, education, mathematics, and psychology. A few research areas have below 3.50 complex nominals per T-unit on average. Computer science, education, engineering, mathematics, and psychology. Most research areas had around 2 verb phrases per T-unit. Mathematics, business economics, education, psychology, and computer science had over 2. All data on particular structures can be found in Appendix 10.

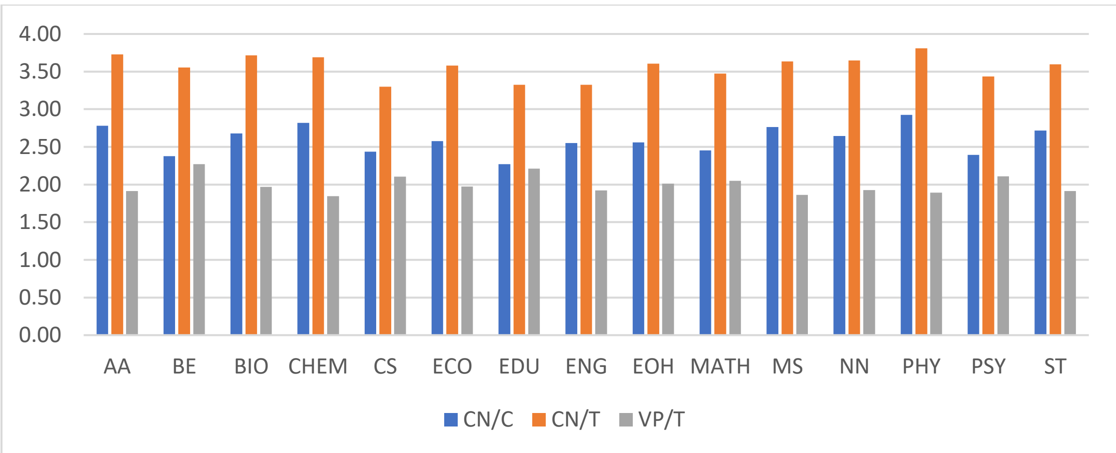

Figure 13. *Particular structures indices across research areas*

5.2. Correlation analysis

Lexical complexity and Altmetric Attention Score

Correlation between the Altmetric Attention Score and lexical complexity indices was low across all research areas. In lexical density the research areas with low positive correlation and statistically significant were biochemistry and biology, ecology, materials science. Mathematics had the highest correlation in Lexical density ($r_s = 0.33$).

Lexical sophistication varied between low positive and low negative correlation. Astrophysics and astronomy, biochemistry and biology, and computer science had a low negative correlation that was statistically significant, whereas mathematics had a negative relationship that was not significant. Education had a low positive correlation, but it was not significant. In verb sophistication, the effects were non-existent in all research areas except mathematics, which showed a weak relationship between verb variation (VV1 $r_s = 0.30$, SVV1 $r_s = 0.30$), which was statistically significant and biochemistry and biology, but it was not statistically significant.

Astrophysics and astronomy had the highest NDW ($r_s = 0.34$), but it was not statistically significant. The only lexical index that had a high positive correlation was the Type-token ratio CTTR in computer science ($r_s = 0.64$), which is visualized in Figure 14. However, it was not statistically significant. Astronomy and astrophysics had a weak positive relationship in CTTR ($r_s = 0.32$) and RTTR ($r_s = 0.32$), but not statistically significant.

Business economics, education, ecology, neuroscience and neurology, physics, psychology, and environmental and occupational health had the lowest correlation between lexical complexity and AAS, which all ranged from weak to no relationship. The correlation coefficients of all research areas can be found in Appendix 11. P-values can be found in Appendix 13.

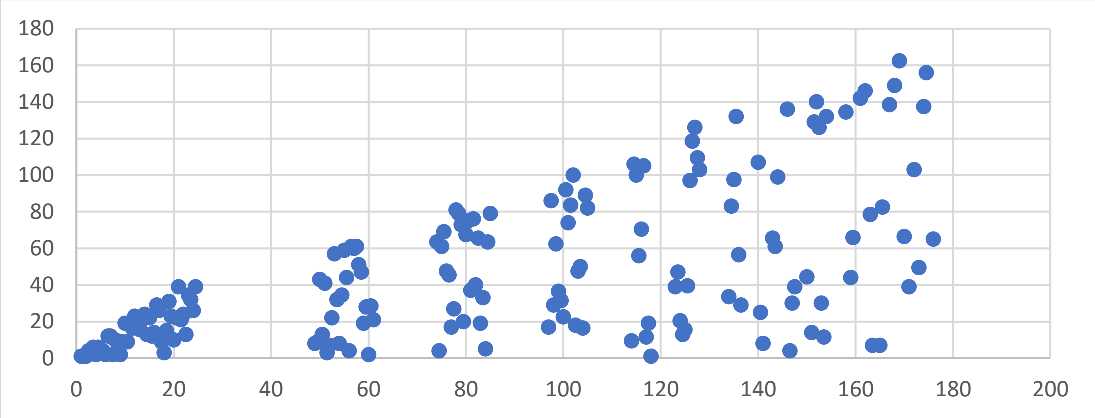

Figure 14. *Visualization of the positive relationship between CTTR and AAS in computer science (r_s = 0.64)*

Syntactic complexity and Altmetric Attention Score

The results found between syntactic complexity and AAS were similar to the results of lexical complexity and AAS. The correlation was low for all indices and across all research areas except computer science. Computer science showed a medium to high positive correlation between all syntactic complexity indices and AAS as shown in Table 13. However, none of them was statistically significant. Complex structures showed the highest correlation(r_s = 0.69), whereas complex nominals per clause had the lowest correlation (r_s = 0.57).

Length of production unit had a positive correlation (MLS r_s = 0.16, MLT r_s = 0.21) in education that was low but statistically significant. The length of clause showed a weak effect on mathematics (MLC r_s = 0.17) but was not statistically significant. Mathematics had a weak negative correlation in Complex structures C/S (r_s = -0.25) and T/S (r_s = -26) that were statistically significant. Business economics also exhibited a low negative correlation with statistical significance in C/S (r_s = -0.12) . Science technology showed the opposite with a low positive correlation (r_s = 0.11).

Verb phrases had a low positive effect on biochemistry and biology, and low negative effect on physics which both were statistically significant. CP/T had a weak correlation in mathematics (r_s = 0.21) that was statistically significant. Similarly, CP/T had a low positive correlation (r_s = 0.24) in mathematics as well. Complex nominals per T-unit had a low positive correlation (r_s = 0.16) in education, and astrophysics and astronomy had a low negative correlation (r_s = -0.13) in complex nominals per clause. Both were statistically significant.

Biochemistry and biology, engineering, environmental and occupational health, and psychology exhibited the lowest correlation between all of the syntactic complexity indices and AAS. The correlation coefficients of all research areas can be found in Appendix 12. P-values can be found in Appendix 14.

Syntactic complexity indices	r_s	*p-value*
MLS	0.67	1.99
MLT	0.65	4.29
MLC	0.62	2.58
C/S	0.69	3.75
VP/T	0.67	2.34
C/T	0.68	1.28
DC/C	0.64	5.55
DC/T	0.65	3.51
T/S	0.68	3.93
CT/T	0.65	3.11
CP/T	0.67	6.15
CP/C	0.66	3.20
CN/T	0.58	2.18
CN/C	0.57	1.14

Table 13. *Correlation between AAS and syntactic complexity in Computer science. Not statistically significant at* $p \leq 0.05$.

6. DISCUSSION

6.1 Lexical and syntactic complexity of abstracts

The first aim of this thesis was to map the lexical and syntactic complexity of abstracts of articles affiliated with Finnish Universities and investigate if there were differences between research areas. The Lexical Complexity Analyser and L2 Syntactic Complexity Analyser was used to assess readability.

Lexical density was similar across all research areas, as were many of the lexical variation indices except number of different words (NDW). Lexical sophistication indices showed more dispersion across research areas. Research areas such as biochemistry and biology, chemistry, materials science, and physics scored the highest on lexical sophistication, which implies heavier use of difficult words in these subjects. Research areas in social sciences such as psychology, education, and business economics scored lower on lexical sophistication. The differences in lexical sophistication can be explained by the use of technical words in the former research areas since technical words are considered part of the complexity of academic language (Warren et al., 2021). The research areas such as neuroscience and neurology, astronomy and astrophysics, ecology, and environmental and occupational health scored high on the level of word variation. Mathematics, education, business economics, and physics had the lowest word variation. Overall, the average number of different words was 110 different words per abstract, and the total number of words per abstract was, on average, 204. Adjective, adverb, and modifier variation were similar, signifying that these aspects of language occur the same amount of time regardless of research area.

Length of production unit exhibited more significant differences between research areas in syntactic complexity than the other complexity indices. On average, astrophysics and astronomy, physics, and mathematics had the longest sentences, T-units, and clauses, whereas Computer science, education, engineering, environmental and occupational health, science technology, and neuroscience and neurology had the shortest length of production unit. Complex structures was highest in business economics and lowest in neuroscience and neurology. However, the difference between research areas was small. Coordination and subordination followed similar patterns across all research areas. The number of dependent clauses was low, as were coordinate clauses. All research areas had over two complex nominals per clause, with physics having almost three complex nominals per clause. Education had the fewest complex nominals. Verb phrases were close to two for all research areas. Business economics had the highest amount of verb phrases, whereas materials science had the fewest.

Results show that research areas are more complex in some indices and less complex in others. No single research area scored the highest in all complexity indices. However, biochemistry and biology was the only research area that, on average, had both higher lexical and syntactic complexity than the average abstract. Research areas with more complex lexicality, such as science technology, chemistry, materials science, and neuroscience and neurology, had lower syntactic complexity. Conversely, research areas with, on average lower lexical complexity, such as psychology, environmental and occupational health, business economics, education, and mathematics, had higher syntactic complexity. The dispersion of data was small, which indicates that the research areas were similarly complex except for some complexity indices such as the number of different words and length of production unit. Syntactic complexity showed more variance across research areas. However, the differences were still small.

The research areas with the highest complexity indices varied across research areas, and to generalise which research areas are the most difficult to read is problematic. Especially since the research areas with high lexical complexity had, lower syntactic complexity and vice versa. Studies using traditional readability formulas can determine which disciplines are the most difficult due to the aggregated score. The results of readability across disciplines by Gazni (2011) are in part supported by the results of this study. Gazni (2011) found that chemistry, medicine, biochemistry, psychology, and social sciences were among the least readable disciplines, whereas mathematics, physics, business economics, and engineering were the easiest to read. Biochemistry and biology and chemistry were, on average, very complex in the results of this thesis; however, they were not the most complex across all indices. Biochemistry and biology scored highest across all lexical sophistication indices, which means that the word difficulty was high. Mathematics had, on average, much lower lexical complexity than other research areas, which can be explained by having the shortest abstracts of all research areas. However, mathematics scored high on multiple syntactic complexity indices, most notably length of production unit. Business economics scored highest on three of the four subordination indices and highest on complex structures, whereas it scored low in many lexical complexity indices. Physics scored highest on complex nominals. The results from the LCA and L2SCA cannot be compared to the

FRE scores used by Gazni (2011) to understand the overall level of difficulty of the abstracts, since the formulas and nature of assessing readability is different.

How complexity in this study scores overall can be compared with Lu's (2012) study on lexical complexity of non-native speaker's oral narratives, Ai and Lu's (2013) and Lu and Ai's (2015) analysis of the syntactic complexity of college-level English essays. The complexity of the studied abstracts and previous studies scored similarly, with a few differences. The abstracts of this study scored on average higher across all lexical complexity indices than in Lu's (2012) study, except for certain lexical variation indices such as NDW, UBER, and verb variation indices. The different styles of texts analysed may explain this. Article abstracts are shorter, whereas Lu (2012) evaluated oral language. Abstracts use fewer different verbs, and the space to express a lot of information means the number of different words are fewer. Lu (2012) also studied the oral language of ESL (English as second language) students, whereas the abstracts are written by scholars and experts, which explains higher lexical sophistication in abstracts. Scholars have more experience in producing academic texts and more knowledge on their chosen topic. Halliday and Matthiessen (2014) note that spoken and written texts are complex in different ways as well. Written language often has more lexical items, which results in more lexically dense texts than spoken language (Halliday & Matthiessen, 2014).

The Syntactic complexity of abstracts scored differently. The length of production unit was higher than Ai and Lu's (2013) and Lu and Ai's (2015) studies of English native speakers' university essays. However, complex structures were slightly lower, which means that the number of clauses per sentence was fewer in research article abstracts than in Ai and Lu (2013) and Lu and Ai (2015). Additionally, the level of subordination (dependent clauses) was lower. However, complex nominals were higher than Ai and Lu (2013) and Lu and Ai (2015) studies. Nominalisation hides agency (Billig, 2008), and turns verb phrases into nouns (Hao & Humphrey, 2019). Fewer verb phrases in the studied abstracts can be explained by the increased use of nominals.

Nominalisation is a common and desired feature of academic language to express concise and precise information (Snow, 2010). Lexical complexity may be higher due to technical language and researchers' knowledge of their respective topics.

Lexical sophistication is another aspect of academic language that contributes to its perceived difficulty (Snow, 2010). Another reason for lexical sophistication being more complex in abstracts is the comparison of written text and spoken language. A comparison between written and spoken language may be insufficient to gauge the overall complexity of texts. However, no similar study on lexical complexity and academic language could be located during the writing of this thesis. Overall, research article abstracts use complex language, consistent with studies on academic language and readability (Gazni, 2011; Yeung et al., 2018; Plavén-Sigray et al., 2017; Smith et al., 2017). The conclusions of complexity in comparison with Lu's (2012), Ai and Lu's (2013) and Lu and Ai's (2015) studies are based on the educational level of the participants of those studies, since university level language is regarded as the most difficult in most readability assessments.

The LCA and L2SCA were efficient tools for language assessment. The results were quickly collected, and no coding experience was needed to use the software since the texts can be submitted without tags or lemmatization. The results are thorough and reveal that the complexity of language is not straightforward since higher scores in lexical complexity do not guarantee higher syntactic complexity. As Graesser et al. (2011) mention, traditional readability formulas cannot tell which specific aspects of the text are difficult, but natural language processing tools offer a deeper understanding. However, there are certain disadvantages to using these tools. Data collection is efficient, but the analysis of the results is more time-consuming. Traditional readability formulas have a fixed score and reading grades, enabling easier analysis of where the text difficulty lies compared to other texts. Mesmer (2008) pointed out that traditional readability formulas have a long history and have been tested for many years. The LCA and L2SCA have primarily been used on university students in educational contexts. More data on the complexity of other styles of texts and the texts of different ages and educational groups would be useful for further comparison. Although these programs were developed to evaluate English learners' skills, the indices measured by the programs are useful for readability as well and are well suited for comparisons between texts in the same data. However, analysing overall readability is difficult since no standard score exists to assess lexical and syntactic complexity.

This study made use of new tools to evaluate academic language and readability. Since few similar studies have assessed the readability of academic texts using programs based on natural language processing, this thesis contributes to that area of research. Additionally, this thesis brings more knowledge on which language features are used in academic abstracts. However, as Janan and Wray (2014) and Bailin and Grafstein (2001) mention, readability must incorporate the reader and the reader's previous knowledge to accurately assess if the text is understood. Academic texts are mainly written with other scholars in mind. However, the focus has increasingly turned to the wider public as readers of scientific output (Warren et al. 2021, Worrall et al., 2020). The abstracts exhibited higher complexity in complex nominals than the student level essays, which suggests that reading these abstracts would prove difficult for a wider readership, since research found that students have difficulties understanding scientific texts due to the use of nominalisations (Hao & Humphrey, 2019). The comparison with university students' language shows that the language of abstracts is complex, but if it is too difficult for a wider readership is only possible to answer by asking the readers themselves.

6.2 Readability and online attention

The second aim of this thesis was to investigate if lexical and syntactic complexity influences the online attention received by research articles. The research questions were: What is the relationship between readability and online attention, and are there differences between research areas? Online attention was measured with the Altmetric attention score, and correlation analysis was used to assess the relationship between the AAS and lexical and syntactic complexity indices.

The results suggest that overall, lexical complexity had a weak relationship with the Altmetric attention score. There were a few exceptions, such as medium correlation in one lexical variation index in computer science. Lexical density had a low positive correlation in mathematics and materials science that was statistically significant. Lexical sophistication showed a low negative correlation with online attention in astrophysics and astronomy, biochemistry and biology, computer science, and mathematics. The effects were small but statistically significant except in mathematics. On the other hand, mathematics had a low positive relationship with verb sophistication. Lexical variation showed a low negative correlation in astronomy and astrophysics and a low positive correlation in mathematics and materials science. The research areas with the lowest correlation across lexical complexity were business economics, education, ecology, neuroscience and neurology, physics, psychology, and environmental and occupational health. The results indicate that lexical complexity had a small effect on the level of online attention received by research and varied by research area. Lexical sophistication, or the number of difficult words, had a slight negative impact on online attention. Lexical variation showed differences in the positive and negative effects among research areas.

Syntactic complexity exhibited similar results. The only research area with a medium to strong positive correlation was computer science, which showed a positive correlation between all syntactic complexity indices and the altmetric attention score. Mathematics and business economics showed a low negative correlation between complex structures and online attention. After computer Science, mathematics showed the most correlation between syntactic complexity with a negative correlation in T-units per sentence and a slight positive correlation in coordination. The length of sentences, clauses and T-units had no effect on online attention except for a low correlation in education and a high correlation in computer science. The research areas with the lowest correlation between syntactic complexity and the Altmetric attention score were biochemistry and biology, engineering, environmental and occupational health, and psychology.

The effect of lexical and syntactic complexity on online attention varied between research areas. Certain research areas exhibited no relationship between complexity indices and the Altmetric attention score, and others showed a weak link. Computer science was the only research area with a medium to strong positive relationship between syntactic complexity and online attention. This implies that syntactically complex abstracts receive more attention. Otherwise, the syntactic complexity exhibited a more negligible influence on the altmetric attention score than lexical complexity.

The results reflect an inconclusive picture of the influence of readability and online attention, similar to results from previous research. Previous research on readability indicates a possible link between the difficulty of texts and the level of scientific impact. Gazni (2011) found that low readability positively affects citations, but Didegah and Thelwall (2013) discovered that it negatively impacts research areas such as biochemistry. Contrastingly, the results in this study showed that biochemistry and biology had a positive relationship between verb phrases and AAS, but the overall effect was small. Lei and Yan (2016) discovered no connection between readability and citations, which were true for several of research areas of the results of this study. The results of this thesis also partly support to Sienkiewicz and Altmann's (2016) study, which found that the effects of readability were small, and the complexity of texts has no overall impact. This thesis provides further evidence that the influence of readability varies between disciplines and research areas.

Jin et al. (2021) found that lexical sophistication negatively correlated with the AAS. The results of this thesis showed similar patterns. However, the effects were minor and only valid for some research areas. On the other hand, verb sophistication showed a weak positive correlation in mathematics, which was statistically significant. Similar findings were not reported in Jin et al. (2021). Lexical variation also showed a positive correlation in Jin et al. (2021). The data in this thesis suggest that lexical variation had a low correlation across all research areas, except for one index in computer science. The results were inconclusive since some research areas reported a negative correlation in some lexical variation indices and others positive. Jin et al. (2021) also found a correlation between verb phrases and complex nominals to influence online attention. Computer science was the only research area with a strong positive relationship in both verb phrases and complex nominals. The results showed that biochemistry and biology had a weak positive relationship, but physics had a weak negative relationship. Chen et al. (2020) found that the sentence length affected downloads, which was not supported by the results in this thesis. Differences between the effects of in thesis and other studies can be explained by sample size, disciplines analysed, and research methods.

The descriptive data on coverage of altmetric sources showed that research received the most visibility on Twitter. However, this fails to prove that non-experts share research articles more. Jin et al. (2021) theorise that lexical sophistication negatively affects online attention due to non-experts sharing articles with less difficult words. There is no indication that non-experts share more research on Twitter in the results of this thesis, and this is something that needs further study. Instead, research has shown that scholars increasingly use social media in scholarly communication (Haustein et al., 2016), and research is predominantly shared by experts (Vainio &

Holmberg, 2017). Additionally, the number of difficult words only had a minor influence on some research areas, which would indicate that the users who shared these articles are less affected by the use of difficult words in research article abstracts. This implies that experts share these research articles, but such conclusions need to be supported by data on Twitter users, which is outside the scope of this study.

Research areas such as computer science and mathematics showed the highest connection between readability and online attention. In computer science, the results imply that more grammatically complex language results in more attention, whereas the influence in mathematics varied. For instance, difficult verbs and the number of different words resulted in more attention. Syntactic complexity, however, resulted in less online attention in mathematics. 60% of all research articles had at least one recorded altmetric event as shown in Table 11, however, computer science and mathematics had 39% and 32%. These two research areas had the lowest attention scores and the least amount of coverage, which may explain the larger impact of lexical and syntactic features. Another important aspect to consider is the way abstracts are written and if that influences attention. Mathematics had on average the shortest abstracts, which may decrease its informativeness and lead to less attention.

A question is whether the nature of the research itself and its affiliation to Finnish universities affect online attention rather than language features. For instance, the article with the highest AAS in neuroscience and neurology, titled "Sauna bathing reduces the risk of stroke in Finnish men and women", suggests that people share articles relevant to their lifestyle and health. This has also been stated by Vainio & Holmberg (2019). However, this conclusion does not apply to all research areas since the most shared article with the highest AAS in the whole dataset was an article from Biochemistry biology, "CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response", which uses technical and field-specific terminology. This could further prove that the articles are primarily shared by experts, and thus, low readability poses little problems. Or the subject matter is again the influencing factor which generates higher attention. Since altmetrics are often highly skewed, the question arises of how the results would differ if outlier like these were removed.

The strength of this study was the sample size and the different research areas included. Previous research has primarily focused on one discipline, or a few select journals. Few studies have been done on lexical and syntactic complexity and the impact on online attention. This thesis contributes to further understanding of what factors influence the online attention of research. However, the results were inconclusive since the effect of lexical and syntactic complexity varied between research areas and the effect was positive or negative depending on the complexity index and research area. More research is needed on readability and its effect on online attention.

6.3 Limitations

This thesis focused on the top 15 research areas with the most published research articles in 2018, however, there was a strong focus on technical research areas, which resulted in a lack of coverage of research in social sciences and humanities. Furthermore, only abstracts from articles written in 2018 were selected for analysis, which means the results cannot be generalised. The results may also have showed greater differences in readability and the effect on online attention if two distinct research areas, such medical sciences and humanities, were compared. Another limitation is the LCA and L2SCA. These programs have not been extensively used, and the lexical complexity results were compared to spoken language. Lastly, the altmetric attention score does not cover all online events and Altmetric explorer does not track all sources of altmetrics. In addition, the Mendeley readership number was not included in this thesis despite social reference managers being extensively used and a significant aspect of altmetrics.

7. FUTURE RESEARCH

The focus of this thesis was the lexical and syntactic complexity of research article abstracts and how this influences the level of online attention research receives. The results indicate that readability affects the online attention of research to a certain degree and depending on discipline. Future research could incorporate more research areas and disciplines in lexical and syntactic analysis, as well as the full-body texts of research articles. Other aspects that need to be considered are differences between social media platforms and other altmetric sources and the level of complexity or whether non-experts share more readable research articles. Another thing to consider is the articles that have received no online attention. What is the readability of these articles, and are there differences between them and articles that have received online attention? Although the effects of readability on online attention were small, further investigation into the subject is needed to gain a more comprehensive understanding of the effects of language features and the impact on research.

REFERENCES

Academy of Finland. (n.d.). *Review and decision-making*. Retrieved December 15, 2021, from https://www.aka.fi/en/research-funding/peer-review-and-funding-decision/review-and-decision-making/

Ai, H. & Lu, X. (2010). A Web-based System for Automatic Measurement of Lexical Complexity. https://doi.org/10.13140/RG.2.2.16499.07208

Ai, H & Lu, X. (2013). A corpus-based comparison of syntactic complexity in NNS and NS university students writing. In A. Díaz-Negrillo, N. Ballier & P. Thompson (Eds.), *Automatic Treatment and Analysis of Learner Corpus Data*, (pp. 249-264). Amsterdam/Philadelphia: John Benjamins.

Alhoori, H. (2019). Anatomy of scholarly information behavior patterns in the wake of academic social media platforms. *International Journal on Digital Libraries* 20, 369–389. https://doi.org/10.1007/s00799-018-0255-9

Almind, T. C. & Ingwersen, P. (1997). Informetric analyses on the world wide web: Methodological approaches to 'webometrics'. *Journal of documentation*, 53(4), 404-426. https://doi.org/10.1108/EUM0000000007205

Altmetric.com. (n.d.a). *How is the Altmetric Attention Score calculated?* Retrieved August 30, 2021, from https://help.altmetric.com/support/solutions/articles/6000233311-how-is-the-altmetric-attention-score-calculated-

Altmetric.com. (n.d.b) *The donut and Altmetric Attention Score*. Retrieved August 30, 2021, from https://www.altmetric.com/about-our-data/the-donut-and-score/

Altmetric.com. (n.d.c). *What are altmetrics?* Retrieved August 30, 2021, from https://www.altmetric.com/about-altmetrics/what-are-altmetrics/

Andrés, A. (2009). *Measuring academic research: How to undertake a bibliometric study*. Chandos Publishing.

Arroyo-Machado, W., Torres-Salinas, D., Herrera-Viedma, E., Romero-Frias, E. & Lozano, S. (2020). Science through Wikipedia: A novel representation of open knowledge through co-citation networks. *PloS one*, 15(2), e0228713. https://doi.org/10.1371/journal.pone.0228713

Bailin, A. & Grafstein, A. (2001). The linguistic assumptions underlying readability formulae: A critique. *Language & communication*, 21(3), 285-301. https://doi.org/10.1016/S0271-5309(01)00005-2

Bar-Ilan, J. (2001). Data collection methods on the Web for informetric purposes: A review and analysis. *Scientometrics*, 50(1), 7-32.

Bar-Ilan, J. (2008). Which h-index? — A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, 74(2), 257-271. https://doi.org/10.1007/s11192-008-0216-y

Bar-Ilan, J., Halevi, G. & Milojević, S. (2019). Differences between Altmetric Data Sources – A Case Study. *The journal of altmetrics*, 2(1), 1. https://doi.org/10.29024/joa.4

Basch, C. H., Mohlman, J., Hillyer, G. C. & Garcia, P. (2020). Public Health Communication in Time of Crisis: Readability of On-Line COVID-19 Information.

Disaster medicine and public health preparedness, 14(5), 635-637. https://doi.org/10.1017/dmp.2020.151

Basch CH, Mohlman J, Hillyer GC, Garcia P. (2021). Simplifying language is inclusive and improves access to up-to-date scientific information. *Disaster medicine and public health preparedness*, 1. https://doi.org/10.1017/dmp.2021.87

Beaudry, C. & Larivière, V. (2016). Which gender gap? Factors affecting researchers' scientific impact in science and medicine. *Research policy*, 45(9), 1790-1817. https://doi.org/10.1016/j.respol.2016.05.009

Begeny, J. C. & Greene, D. J. (2014). Can readability formulas be used to successfully gauge difficulty of reading materials? *Psychology in the schools*, 51(2), 198-215. https://doi.org/10.1002/pits.21740

Benjamin, R. G. (2012). Reconstructing Readability: Recent Developments and Recommendations in the Analysis of Text Difficulty. *Educational psychology review*, 24(1), 63-88. https://doi.org/10.1007/s10648-011-9181-8

Bhattacherjee, A. (2012). *Social science research*. University of South Florida. https://digitalcommons.usf.edu/cgi/viewcontent.cgi?article=1002&context=oa_textbooks

Billig, M. (2008). The language of critical discourse analysis: The case of nominalization. *Discourse & society, 19*(6), 783-800. https://doi.org/10.1177/0957926508095894

Björk, B. & Solomon, D. (2012). Open access versus subscription journals: A comparison of scientific impact. *BMC medicine*, 10(1), 73. https://doi.org/10.1186/1741-7015-10-73

Björneborn, L. & Ingwersen, P. (2001). Perspectives of webometrics. *Scientometrics*, 50(1), 65-82.

Björneborn, L. & Ingwersen, P. (2004). Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology*, 55(14), 1216-1227. https://doi.org/10.1002/asi.20077

Bloch, S. & Walter, G. (2001). The Impact Factor: Time for change. *Australian and New Zealand journal of psychiatry*, 35(5), 563-568. https://doi.org/10.1080/0004867010060502

Borgman, C. & Furner, J. (2005). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology*, 36(1), 2-7.

Bornmann, L. (2012). Measuring the societal impact of research: Research is less and less assessed on scientific impact alone--we should aim to quantify the increasingly important contributions of science to society. *EMBO reports*, 13(8), 673-676. https://doi.org/10.1038/embor.2012.99

Bornmann, L. (2015). Alternative metrics in scientometrics: A meta-analysis of research into three altmetrics. *Scientometrics*, 103(3), 1123-1144. https://doi.org/10.1007/s11192-015-1565-y

Bornmann, L. & Daniel, H. (2009). The state of h index research. Is the h index the ideal way to measure research performance? *EMBO reports*, 10(1), 2-6. https://doi.org/10.1038/embor.2008.233

Bornmann, L. & Leydesdorff, L. (2014). Scientometrics in a changing research landscape. *EMBO reports*, 15(12), 1228-1232. https://doi.org/10.15252/embr.201439608

Bornmann, L. & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, 66(11), 2215-2222. https://doi.org/10.1002/asi.23329

Broadus, R. N. (1987). Toward a definition of bibliometrics. *Scientometrics*, 12(5-6), 373-379. https://doi.org/10.1007/BF02016680

Brown, B. A., Donovan, B. & Wild, A. (2019). Language and cognitive interference: How using complex scientific language limits cognitive performance. *Science education,* 103(4), 750-769. https://doi.org/10.1002/sce.21509

Chen, B., Deng, D., Zhong, Z. & Zhang, C. (2020). Exploring linguistic characteristics of highly browsed and downloaded academic articles. *Scientometrics*, 122(3), 1769-1790. https://doi.org/10.1007/s11192-020-03361-4

Chen, P., Hayes, E., Larivière, V. & Sugimoto, C. R. (2018). Social reference managers and their users: A survey of demographics and ideologies. *PloS one*, 13(7), e0198033. https://doi.org/10.1371/journal.pone.0198033

Cohen, K. B., Johnson, H. L., Verspoor, K., Roeder, C. & Hunter, L. E. (2010). The structural and content aspects of abstracts versus bodies of full text journal articles are different. *BMC bioinformatics*, 11(1), 492. https://doi.org/10.1186/1471-2105-11-492

Costas, R., Zahedi, Z. & Wouters, P. (2015). Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, 66(10), 2003-2019. https://doi.org/10.1002/asi.23309

Crossley, S. A., Allen, D. B. & McNamara, D. S. (2011). Text Readability and Intuitive Simplification: A Comparison of Readability Formulas. *Reading in a foreign language*, 23(1), 84.

Crossley, S. A., Greenfield, J. & McNamara, D. S. (2008). Assessing Text Readability Using Cognitively Based Indices. *TESOL quarterly*, 42(3), 475-493. https://doi.org/10.1002/j.1545-7249.2008.tb00142.x

Crossley, S. A., Skalicky, S. & Dascalu, M. (2019). Moving beyond classic readability formulas: New methods and new models. *Journal of research in reading*, 42(3-4), 541-561. https://doi.org/10.1111/1467-9817.12283

Dale, E. & Chall, J. S. (1948). A Formula for Predicting Readability. *Educational research bulletin*, 27(1), 11-28.

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

Daraio, C. (2021). Altmetrics as an Answer to the Need for Democratization of Research and Its Evaluation. *The journal of altmetrics*, 4(1). https://doi.org/10.29024/joa.43

Didegah, F. & Thelwall, M. (2013). Which factors help authors produce the highest impact research? Collaboration, journal and document properties. *Journal of informetrics*, 7(4), 861-873. https://doi.org/10.1016/j.joi.2013.08.006

Dolnicar, S., & Chapple, A. (2015). The readability of articles in tourism journals. *Annals of Tourism Research*, 52, 161–166.

DOAJ. (n.d.). About DOAJ. Retrieved February 30, 2022, from https://doaj.org/about/

DORA. (n.d.). The San Francisco Declaration of Research Assessment. Retrieved February 30, 2022, from https://sfdora.org/read/

DuBay, W. H. (2004). *The Principles of Readability*. https://www.researchgate.net/publication/228965813_The_Principles_of_Readability

Dziubaniuk, O., Barner-Rasmussen, W., Koporcic, N., Ivanova-Gongne, M., Mandják, T. & Markovic, S. (2021). Business-to-business marketing research: Assessing readability and discussing relevance to practitioners. *Industrial marketing management*, 92, 217-231. https://doi.org/10.1016/j.indmarman.2020.01.012

Egghe, L. (2005). Expansion of the field of informetrics: Origins and consequences. *Information processing & management*, 41(6), 1311-1316. https://doi.org/10.1016/j.ipm.2005.03.011

Eysenbach, G. (2011). Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *Journal of medical Internet research*, 13(4), e123. https://doi.org/10.2196/jmir.2012

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221–233. https://doi.org/10.1037/h0057532

Fraumann, G. (2018). The Values and Limits of Altmetrics. *New directions for institutional research*, 2018(178), 53-69. https://doi.org/10.1002/ir.20267

García-Villar, C. (2021). A critical review on altmetrics: Can we measure the social impact factor? *Insights into imaging*, 12(1), 92. https://doi.org/10.1186/s13244-021-01033-2

Garfield, E. (1973). Citation frequency as a measure of research activity and performance. *Essays of an Information Scientist*,1, 406-408. http://www.garfield.library.upenn.edu/essays/V1p406y1962-73.pdf

Garfield, E. & Sher, I. H. (1963). New factors in the evaluation of scientific literature through citation indexing. *American documentation*, 14(3), 195-201. https://doi.org/10.1002/asi.5090140304

Gazni, A. (2011). Are the abstracts of high impact articles more readable? Investigating the evidence from top research institutions in the world. *Journal of information science*, 37(3), 273-281. https://doi.org/10.1177/0165551511401658

Gilliver, S. (2015). Plain language and readability. *Medical writing*, 24(1), 1-2. https://doi.org/10.1179/2047480614Z.000000000258

Graesser, A. C., McNamara, D. S. & Kulikowich, J. M. (2011). Coh-Metrix: Providing Multilevel Analyses of Text Characteristics. *Educational researcher*, 40(5), 223-234. https://doi.org/10.3102/0013189X11413260

Green, T. (2019). Is open access affordable? Why current models do not work and why we need internet-era transformation of scholarly communications. *Learned publishing*, 32(1), 13-25. https://doi.org/10.1002/leap.1219

Guerini, M., Pepe, A. & Lepri, B. (2012). Do Linguistic Style and Readability of Scientific Abstracts affect their Virality? https://doi.org/10.48550/arXiv.1203.4238

Gunning, R. (1952). *The Technique of Clear Writing*. McGraw-Hill.

Haas, K., Brillante, C., Sharp, L., Elzokaky, A. K., Pasquinelli, M., Feldman, L., Kovitz, K., Joo, M. (2018). Lung cancer screening: Assessment of health literacy and readability of online educational resources. *BMC public health, 18*(1), 1356. https://doi.org/10.1186/s12889-018-6278-8

Halliday, M. A. K. & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar* (4th ed.). Routledge.

Hammarfelt, B. (2014). Using altmetrics for assessing research impact in the humanities. *Scientometrics*, 101(2), 1419-1430. https://doi.org/10.1007/s11192-014-1261-3

Hao, J. & Humphrey, S. L. (2019). Reading nominalizations in senior science. *Journal of English for academic purposes*, 42, 100793. https://doi.org/10.1016/j.jeap.2019.100793

Hartley, J. (2004). Current findings from research on structured abstracts. *Journal of the Medical Library Association*, 92(3), 368-371.

Hartley, J. (2008). *Academic Writing and Publishing: A Practical Handbook* (1st ed.). Routledge. https://doi-org.ezproxy.vasa.abo.fi/10.4324/9780203927984

Hartley, J. (2016). Is time up for the Flesch measure of reading ease? *Scientometrics*, 107(3), 1523-1526.

Haustein, S. (2016) Grand challenges in altmetrics: heterogeneity, data quality and dependencies. *Scientometrics*. 108:413–423. https://doi.org/10.1007/s11192-016-1910-9

Haustein, S., Bowman, T. D., and Costas, R. (2016). Interpreting 'Altmetrics': Viewing Acts on Social Media through the Lens of Citation and Social Theories. In C. Sugimoto (Ed.), *Theories of Informetrics and Scholarly Communication*. Berlin, Boston: De Gruyter Saur. https://doi.org/10.1515/9783110308464

Haustein, S., Peters, I., Bar-Ilan, J., Priem, J., Shema, H. & Terliesner, J. (2014). Coverage and adoption of altmetrics sources in the bibliometric community. *Scientometrics*, 101(2), 1145-1163. https://doi.org/10.1007/s11192-013-1221-3

Haustein, S., Sugimoto, C. R. & Larivière, V. (2015). *Social media in scholarly communication*. https://doi.org/10.1108/AJIM-03-2015-0047

Hicks, D., Wouters, P., Waltman, L. et al. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature,* 520, 429–431. https://doi.org/10.1038/520429a

Hirsch J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569–16572. https://doi.org/10.1073/pnas.0507655102

Holmberg, K. (2015). *Altmetrics for information professionals : Past, present and future*. Elsevier Science & Technology.

Holmberg, K., Bowman, S., Bowman, T., Didegah, F., & Kortelainen, T. (2019a). What Is Societal Impact and Where Do Altmetrics Fit into the Equation?. *Journal of Altmetrics*, 2(1), 6. http://doi.org/10.29024/joa.21

Holmberg, K., Bowman, T., Didegah, F., & Lehtimäki, J. (2019b). The Relationship Between Institutional Factors, Citation and Altmetric Counts of Publications from Finnish Universities. *Journal of Altmetrics*, 2(1), 5. http://doi.org/10.29024/joa.20

Holmberg, K., Hedman, J., Bowman, T. D., Didegah, F. & Laakso, M. (2020). Do articles in open access journals have more frequent altmetric activity than articles in subscription-based journals? An investigation of the research output of Finnish universities. *Scientometrics*, 122(1), 645-659. https://doi.org/10.1007/s11192-019-03301-x

Holmberg, K. & Park, H. W. (2018). An altmetric investigation of the online visibility of South Korea-based scientific journals. *Scientometrics*, 117(1), 603-613. https://doi.org/10.1007/s11192-018-2874-8

Holmberg, K. & Thelwall, M. (2014). Disciplinary differences in Twitter scholarly communication. *Scientometrics, 101*(2), 1027-1042. https://doi.org/10.1007/s11192-014-1229-3

Hood, W. W. & Wilson, C. S. (2001). The Literature of Bibliometrics, Scientometrics, and Informetrics. *Scientometrics*, 52(2), 291-314. https://doi.org/10.1023/A:1017919924342

Hosseini, M-S & Akbarzadeh, MA. (2021). Persisting on readability could provoke the risk of misinformation: A COVID-19 pandemic concern. *Disaster medicine and public health preparedness*, 1. https://doi.org/10.1017/dmp.2021.25

Houghton, J. W., Steele, C. & Henty, M. (2004). Research practices and scholarly communication in the digital environment. *Learned publishing*, 17(3), 231-249. https://doi.org/10.1087/095315104323159667

Ingwersen, P. (1998). The calculation of web impact factors. *Journal of documentation*, 54(2), 236-243. https://doi.org/10.1108/EUM0000000007167

Janan, D., & Wray, D. (2014). Reassessing the accuracy and use of readability formulae. *Malaysian journal of learning and instruction*, 11, 127-145. https://doi.org/10.32890/mjli.11.2014.7668

Jin, T., Duan , H., Lu, X., Ni, J. & Guo, K. (2021). Do research articles with more readable abstracts receive higher online attention? Evidence from Science. *Scientometrics*, 126(10), 8471. https://doi.org/10.1007/s11192-021-04112-9

Jubb, M. (2013). Introduction: Scholarly communications – disruptions in a complex ecology. In D. Shorley & M. Jubb (Eds.). *The future of scholarly communication*. Facet Publishing.

Just, M. A. & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review*, 87(4), 329-354. https://doi.org/10.1037/0033-295X.87.4.329

Karmakar, M., Banshal, S. K. & Singh, V. K. (2021). A large-scale comparison of coverage and mentions captured by the two altmetric aggregators: Altmetric.com and PlumX. *Scientometrics*, 126(5), 4465-4489. https://doi.org/10.1007/s11192-021-03941-y

King, R. (1976). A comparison of the readability of abstracts with their source documents. *Journal of the American Society for Information Science*, 27(2), 118-121. https://doi.org/10.1002/asi.4630270207

Klare, G. (1963). *The measurement of reading*. Iowa State University Press.

Klare, G. R. (1974). Assessing Readability. *Reading research quarterly*, 10(1), 62-102. https://doi.org/10.2307/747086

Kothari, C. R. (2004). *Research methodology: Methods & techniques* (2nd rev. ed.). New Age International (P) Ltd., Publishers.

Larivière, V., Haustein, S. & Börner, K. (2015). Long-distance interdisciplinarity leads to higher scientific impact. *PloS one*, 10(3), e0122565. https://doi.org/10.1371/journal.pone.0122565

Lee, S. & French, N. (2011). The readability of academic papers in the Journal of Property Investment & Finance. *Journal of property investment & finance*, 29(6), 693-704. https://doi.org/10.1108/14635781111171814

Lei, L. & Yan, S. (2016). Readability and citations in information science: Evidence from abstracts and articles of four journals (2003–2012). *Scientometrics*, 108(3), 1155-1169. https://doi.org/10.1007/s11192-016-2036-9

Leydesdorff, L., Bornmann, L. & Adams, J. (2019). The integrated impact indicator revisited (I3): A non-parametric alternative to the journal impact factor. *Scientometrics*, 119(3), 1669-1694. https://doi.org/10.1007/s11192-019-03099-8

Lu, X. (2010). Automatic Analysis of Syntactic Complexity in Second Language Writing. *International Journal of Corpus Linguistics,* 15(4), 474-496. https://doi.org/10.1075/ijcl.15.4.02lu

Lu, X. (2012). The Relationship of Lexical Richness to the Quality of ESL Learners' Oral Narratives. *The Modern language journal (Boulder, Colo.),* 96(2), 190-208. https://doi.org/10.1111/j.1540-4781.2011.01232.x

Lu, X & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing*, 29, 16-27.

Lu, X., & Bluemel, B. (2020). Automated assessment of language. In S. Conrad, A. Hartig, & L. Santelmann (Eds.), *The Cambridge introduction to applied linguistics* (pp. 86-93). Cambridge: Cambridge University Press.

Maamuujav, U., Olson, C. B. & Chung, H. (2021). Syntactic and lexical features of adolescent L2 students' academic writing. *Journal of second language writing, 53*, 100822. https://doi.org/10.1016/j.jslw.2021.100822

Maass, C. (2020). *Easy language – plain language – easy language plus: Balancing comprehensibility and acceptability*. Frank & Timme.

Manning, C. D., Prabhakar Raghavan, P., and Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.

McCannon, B. C. (2019). Readability and research impact. *Economics letters, 180*, 76-79. https://doi.org/10.1016/j.econlet.2019.02.017

McLaughlin, G. H. M. (1969). SMOG Grading-a New Readability Formula. *Journal of reading*, 12(8), 639-646.

McLaughlin, G. H. (1974). Temptations of the Flesch. *Instructional science*, 2(4), 367-383. https://doi.org/10.1007/BF00123459

Merton, R.K. (1968), The Matthew effect in science, *Science*, 159, 56-63.

Merton, R.K. (Ed.). (1973). *The Sociology of Science: Theoretical and Empirical Investigations*. University of Chicago Press, Chicago, IL

Mesmer, H. A. E. (2008). *Tools for matching readers to texts: Research-based practices*. Guilford Press.

Mingers, J. & Leydesdorff, L. (2015). A review of theory and practice in scientometrics. *European journal of operational research*, 246(1), 1-19. https://doi.org/10.1016/j.ejor.2015.04.002

Morrison, H. (2009). *Scholarly communication for librarians*. Chandos Publishing.

Ortega, J. L. (2018). Reliability and accuracy of altmetric providers: A comparison among Altmetric.com, PlumX and Crossref Event Data. *Scientometrics, 116*(3), 2123-2138. https://doi.org/10.1007/s11192-018-2838-z

Ortega, J. L. (2019). Blogs and news sources coverage in altmetrics data providers: A comparative analysis by country, language, and subject. *Scientometrics*, 122(1), 555-572. https://doi.org/10.1007/s11192-019-03299-2

Osborne, R. (2015). Open Access Publishing, academic research and scholarly communication. *Online information review*, 39(5), 637-648. https://doi.org/10.1108/OIR-03-2015-0083

Pagell (2014). Bibliometrics and University Research Rankings Demystified for Librarians. In C. Chen & R. Larsen (Eds.). *Library and Information Sciences: Trends and Research* (1st ed). Springer Berlin Heidelberg.

Perfetti, C. (2007). Reading Ability: Lexical Quality to Comprehension. *Scientific studies of reading*, 11(4), 357-383. https://doi.org/10.1080/10888430701530730

Plavén-Sigray, P., Matheson, G. J., Schiffler, B. C. & Thompson, W. H. (2017). The readability of scientific texts is decreasing over time. *eLife, 6*, e27725. https://doi.org/10.7554/eLife.27725

Plum Analytics .(n.d.a). *About Plum Metrics*. Retrieved February 13, 2022, from https://plumanalytics.com/learn/about-metrics/

Plum Analytics. (n.d.b). *About Artifacts*. Retrieved February 13, 2022, from https://plumanalytics.com/learn/about-artifacts/

Pooladian, A. & Borrego, Á. (2017). Methodological issues in measuring citations in Wikipedia: A case study in Library and Information Science. *Scientometrics*, 113(1), 455-464. https://doi.org/10.1007/s11192-017-2474-z

Priem, J., Groth, P. & Taraborelli, D. (2012). The Altmetrics Collection. *PloS one*, 7(11), e48753. https://doi.org/10.1371/journal.pone.0048753

Priem, J., Taraborelli, D., Groth, P. & Neylon, C. (2010), *Altmetrics: A manifesto*, 26 October 2010. http://altmetrics.org/manifesto

Pritchard, A. (1969). Statistical bibliography or bibliometrics. *Journal of documentation,* 25(4), 348-349.

Pulido, C. M., Redondo-Sama, G., Sordé-Martí, T. & Flecha, R. (2018). Social impact in social media: A new method to evaluate the social impact of research. *PloS one*, 13(8), e0203117. https://doi.org/10.1371/journal.pone.0203117

Richardson, M. A., Bernstein, D. N. & Mesfin, A. (2021). Manuscript characteristics associated with the altmetrics score and social media presence: An analysis of seven spine journals. *The spine journal*, 21(4), 548-554. https://doi.org/10.1016/j.spinee.2020.11.001

Ruano, J., Aguilar-Luque, M., Isla-Tejera, B., Alcalde-Mellado, P., Gay-Mimbrera, J., Hernández-Romero, J. L., Sanz-Cabanillas, J. L., Maestre-López, B., González-Padilla, M., Carmona-Fernández, P. J., Gómez-García, F., & García-Nieto, A. V. (2018). Relationships between abstract features and methodological quality explained variations of social media activity derived from systematic reviews about psoriasis interventions. *Journal of clinical epidemiology*, 101, 35–43. https://doi.org/10.1016/j.jclinepi.2018.05.015

Sawyer, A. G., Laran, J. & Xu, J. (2008). The Readability of Marketing Journals: Are Award-Winning Articles Better Written? *Journal of marketing*, 72(1), 108-117. https://doi.org/10.1509/jmkg.72.1.108

Schriver, K. A. (2017). Plain Language in the US Gains Momentum: 1940-2015. *IEEE transactions on professional communication*, 60(4), 343-383. https://doi.org/10.1109/TPC.2017.2765118

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ*, 314(7079), 497-502. https://doi.org/10.1136/bmj.314.7079.497

Segre, S. (2014). *Contemporary sociological thinkers and theories*. Ashgate Publishing.

Sheedy, J. E., Subbaram, M. V., Zimmerman, A. B. & Hayes, J. R. (2005). Text Legibility and the Letter Superiority Effect. *Human factors*, 47(4), 797-815. https://doi.org/10.1518/001872005775570998

Shehata, A., Ellis, D. & Foster, A. (2015). Scholarly communication trends in the digital age. *Electronic library*, 33(6), 1150-1162. https://doi.org/10.1108/EL-09-2014-0160

Shema, H., Bar-Ilan, J. & Thelwall, M. (2012). Research blogs and the discussion of scholarly information. *PloS one*, 7(5), e35869. https://doi.org/10.1371/journal.pone.0035869

Shema, H., Bar-Ilan, J. & Thelwall, M. (2014). Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics. *Journal of the Association for Information Science and Technology*, 65(5), 1018-1027. https://doi.org/10.1002/asi.23037

Sienkiewicz, J. & Altmann, E. G. (2016). Impact of lexical and sentiment factors on the popularity of scientific papers. *Royal Society open science*, 3(6), 160140. https://doi.org/10.1098/rsos.160140

Smith, K., Buchanan, P. & McDonald, P. (2017). How easy is it for a lay audience to read medical journals? A survey of the readability scores of a sample of research papers on diabetes. *The Lancet (British edition),* 390, S82. https://doi.org/10.1016/S0140-6736(17)33017-9

Snow, C. E. (2010). Academic Language and the Challenge of Reading for Learning About Science. *Science (American Association for the Advancement of Science*), 328(5977), 450-452. https://doi.org/10.1126/science.1182597

Stanford NLP Group. (n.d.). Software > Stanford Log-linear Part-Of-Speech Tagger. Retrieved March 29, 2022, from https://nlp.stanford.edu/software/tagger.shtml

Statology. (2020). How to Calculate Spearman Rank Correlation in Excel. Retrieved February 12, 2022, from https://www.statology.org/spearman-rank-correlation-excel/

Stricker, J., Chasiotis, A., Kerwer, M. & Günther, A. (2020). Scientific abstracts and plain language summaries in psychology: A comparison based on readability indices. *PloS one*, 15(4), e0231160. https://doi.org/10.1371/journal.pone.0231160

Sud, P. & Thelwall, M. (2013). Evaluating altmetrics. *Scientometrics*, 98(2), 1131-1143. https://doi.org/10.1007/s11192-013-1117-2

Sugimoto, C.R. (2015). “Attention is not impact” and other challenges for altmetrics. Wiley Exch. Retrieved February 20, 2022, from https://www.wiley.com/network/researchers/promoting-your-article/attention-is-not-impact-and-other-challenges-for-altmetrics

Tague-Sutcliffe, J. (1992). An introduction to informetrics. *Information Processing and Management,* 28(1), 1-3.

Thelwall, M. (2008). Bibliometrics to webometrics. *Journal of Information Science*, 34 (4) 2008, 605–621. https://doi.org/10.1177/0165551507087238

Thelwall, M. (2018). Early Mendeley readers correlate with later citation counts. *Scientometrics*, 115(3), 1231-1240. https://doi.org/10.1007/s11192-018-2715-9

Thelwall, M. (2020). The Pros and Cons of the Use of Altmetrics in Research Assessment. *Scholarly assessment reports*, 2(1), 2. https://doi.org/10.29024/sar.10

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C.R. (2013). Do altmetrics work? Twitter and ten other social web services. *PLoS ONE*, 8(5). https://doi:10.1371/journal.pone.0064841

Vainio, J. & Holmberg, K. (2017). Highly tweeted science articles: Who tweets them? An analysis of Twitter user profile descriptions. *Scientometrics*, 112(1), 345-366. https://doi.org/10.1007/s11192-017-2368-0

Vajjala, S., & Meurers, D. (2012). On improving the accuracy of readability classification using insights from second language acquisition. In *Proceedings of the 7th Workshop on Innovative Use of NLP for Building Educational Application*s (pp. 163–173). Montréal, Canada: Association for Computational Linguistics.

Warren, N. L., Farmer, M., Gu, T. & Warren, C. (2021). Marketing Ideas: How to Write Research Articles that Readers Understand and Cite. *Journal of marketing,* 85(5), 42-57. https://doi.org/10.1177/00222429211003560

Weil, B. H. (1970). Standards for writing abstracts. *Journal of the American Society for Information Science*, 21(5), 351–357.

Weimer, K. H. & Andrew, P. G. (2013). How We Participate in the Scholarly Communication Life Cycle. *Journal of map & geography libraries*, 9(3), 217-219. https://doi.org/10.1080/15420353.2013.824397

Ventola, E. (1996). Packing and Unpacking of Information in Academic Texts. In Ventola, E., & Mauranen, A. (Eds.). (1996). *Academic writing : Intercultural and textual issues*. John Benjamins Publishing Company.

Wikipedia. (n.d). *Wikipedia: About*. Retrieved December 8, 2021, from https://en.wikipedia.org/wiki/Wikipedia:About

Worrall, A. P., Connolly, M. J., O'Neill, A., O'Doherty, M., Thornton, K. P., McNally, C., McConkey, S. J., de Barra, E. (2020). Readability of Online COVID-19 Health Information: A Comparison between Four English Speaking Countries. *BMC Public Health*. https://doi.org/10.21203/rs.3.rs-30124/v3

Yeung, A. W. K., Goto, T. K. & Leung, W. K. (2018). Readability of the 100 Most-Cited Neuroimaging Papers Assessed by Common Readability Formulae. *Frontiers in human neuroscience*, 12, 308. https://doi.org/10.3389/fnhum.2018.00308

Yiotis, K. (2005). The open access initiative: A new paradigm for scholarly communications. *Information technology and libraries,* 24(4), 157-162. https://doi.org/10.6017/ital.v24i4.3378

Zahedi, Z. & Van Eck, N. J. (2018). Exploring Topics of Interest of Mendeley Users. *The journal of altmetrics*, 1(1), 5. https://doi.org/10.29024/joa.7

Zhang, L. & Wang, J. (2021). What affects publications' popularity on Twitter? *Scientometrics*, 126(11), 9185-9198. https://doi.org/10.1007/s11192-021-04152-1

APPENDICES

Indices	LD		LS1		LS2		VS1		VS2		CVS1	
R. areas	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*
AA	0.56	0.04	0.55	0.07	0.41	0.05	0.23	0.11	1.23	1.15	0.71	0.34
BE	0.60	0.04	0.38	0.08	0.30	0.06	0.20	0.10	0.90	0.86	0.59	0.31
BIO	0.61	0.04	0.60	0.09	0.46	0.07	0.25	0.11	1.44	1.15	0.78	0.34
CHEM	0.59	0.05	0.60	0.09	0.46	0.07	0.24	0.12	1.14	1.02	0.67	0.34
CS	0.61	0.04	0.44	0.09	0.35	0.07	0.20	0.10	1.08	0.96	0.66	0.33
ECO	0.60	0.04	0.48	0.10	0.37	0.07	0.21	0.10	1.22	1.01	0.71	0.32
EDU	0.59	0.04	0.36	0.08	0.29	0.06	0.19	0.09	0.84	0.77	0.58	0.29
ENG	0.60	0.04	0.49	0.10	0.38	0.07	0.21	0.10	1.09	0.95	0.67	0.31
EOH	0.58	0.04	0.46	0.09	0.33	0.06	0.18	0.09	0.85	0.76	0.58	0.30
MS	0.60	0.04	0.57	0.08	0.44	0.06	0.24	0.11	1.18	1.01	0.70	0.32
MATH	0.57	0.05	0.53	0.10	0.39	0.07	0.20	0.12	0.69	0.74	0.50	0.32
NN	0.60	0.04	0.56	0.09	0.41	0.07	0.21	0.10	1.20	1.06	0.70	0.33
PHY	0.58	0.05	0.57	0.08	0.43	0.07	0.24	0.12	1.09	0.98	0.66	0.33
PSY	0.59	0.04	0.44	0.09	0.33	0.06	0.19	0.09	0.85	0.77	0.59	0.28
ST	0.60	0.04	0.54	0.11	0.41	0.08	0.23	0.11	1.24	1.09	0.71	0.33
ALL	0.59	0.05	0.51	0.12	0.39	0.09	0.22	0.11	1.11	0.99	0.67	0.33

***Appendix 1**. Lexical density and lexical sophistication*

Indices	NDW		NDWZ		NDWERZ		NDWESZ	
R. areas	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*
AA	126.27	36.76	39.16	3.51	38.81	2.07	38.56	2.33
BE	97.12	26.27	38.70	3.19	38.44	2.21	38.68	2.34
BIO	119.74	25.24	39.95	2.81	39.62	1.91	39.30	2.30
CHEM	107.11	27.94	39.38	3.56	39.08	2.25	38.79	2.78
CS	109.20	28.92	39.23	3.21	39.22	2.18	38.90	2.53
ECO	127.56	29.14	39.43	3.12	39.21	1.92	38.90	2.22
EDU	99.49	25.71	38.78	3.05	38.38	1.91	38.32	2.19
ENG	109.94	26.88	39.17	3.12	38.73	2.15	38.43	2.42
EOH	126.78	28.52	39.15	3.04	38.79	1.82	38.23	2.02
MS	105.28	25.18	39.62	3.09	39.00	2.19	38.74	2.41
MATH	78.30	26.09	37.52	3.72	37.62	2.67	37.17	3.15
NN	129.73	25.96	39.75	2.87	39.37	1.86	38.89	2.24
PHY	99.99	27.67	39.04	3.46	38.57	2.25	38.47	2.61
PSY	107.04	24.33	39.14	3.23	38.64	2.06	38.26	2.40
ST	113.56	25.54	40.08	3.07	39.44	1.96	39.28	2.38
ALL	110.63	29.76	39.30	3.25	38.92	2.15	38.69	2.48

***Appendix 2**. Lexical variation – Number of different words*

Indices	TTR		MSTTR		CTTR		RTTR		LOGTTR		UBER	
R. areas	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*
AA	0.54	0.07	0.77	0.04	5.74	0.71	8.11	1.01	0.89	0.02	21.09	3.29
BE	0.58	0.08	0.77	0.05	5.21	0.62	7.37	0.87	0.89	0.02	21.11	4.14
BIO	0.57	0.07	0.79	0.04	5.78	0.58	8.17	0.81	0.89	0.02	22.36	3.42
CHEM	0.59	0.08	0.78	0.05	5.52	0.64	7.80	0.90	0.90	0.02	22.62	4.83
CS	0.57	0.08	0.78	0.05	5.51	0.69	7.80	0.97	0.89	0.02	21.85	3.99
ECO	0.54	0.07	0.78	0.04	5.79	0.62	8.20	0.88	0.89	0.02	21.17	3.25
EDU	0.56	0.07	0.77	0.04	5.23	0.59	7.39	0.84	0.89	0.02	20.47	3.33
ENG	0.56	0.07	0.77	0.05	5.48	0.64	7.75	0.91	0.89	0.02	21.24	3.90
EOH	0.51	0.06	0.77	0.04	5.64	0.63	7.98	0.90	0.88	0.02	19.77	2.66
MS	0.58	0.07	0.78	0.04	5.47	0.61	7.73	0.86	0.89	0.02	22.10	4.14
MATH	0.60	0.10	0.75	0.06	4.71	0.70	6.67	0.98	0.89	0.03	20.36	4.43
NN	0.53	0.07	0.78	0.04	5.79	0.59	8.18	0.84	0.88	0.02	20.77	3.30
PHY	0.58	0.08	0.77	0.05	5.30	0.63	7.49	0.88	0.89	0.02	21.55	4.08
PSY	0.54	0.07	0.77	0.04	5.33	0.62	7.54	0.88	0.88	0.02	20.07	3.30
ST	0.57	0.07	0.79	0.05	5.64	0.57	7.98	0.81	0.89	0.02	22.29	3.92
ALL	0.56	0.08	0.78	0.05	5.50	0.67	7.78	0.94	0.89	0.02	21.47	3.94

Appendix 3*. Lexical variation - Type token ratio and UBER index*

Indices	*LV*		*VV1*		*SVV1*		*CVV1*		*VV2*	
R.areas	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*
AA	0.86	0.10	14.46	5.05	2.65	0.47	0.71	0.08	0.13	0.03
BE	0.87	0.09	14.25	5.29	2.62	0.50	0.71	0.09	0.16	0.04
BIO	0.85	0.10	14.25	4.41	2.64	0.42	0.71	0.08	0.13	0.03
CHEM	0.88	0.10	12.45	4.48	2.45	0.45	0.75	0.09	0.13	0.03
CS	0.84	0.10	15.27	5.23	2.72	0.48	0.71	0.09	0.16	0.03
ECO	0.84	0.10	15.86	4.79	2.78	0.43	0.67	0.08	0.13	0.03
EDU	0.86	0.10	14.14	4.70	2.62	0.45	0.70	0.09	0.16	0.04
ENG	0.84	0.10	14.42	4.62	2.65	0.43	0.70	0.08	0.15	0.03
EOH	0.81	0.12	13.57	4.55	2.57	0.44	0.62	0.08	0.12	0.03
MS	0.86	0.10	12.76	4.17	2.49	0.41	0.73	0.08	0.14	0.03
MATH	0.88	0.11	9.58	4.57	2.13	0.52	0.74	0.10	0.14	0.04
NN	0.81	0.10	14.08	4.61	2.62	0.43	0.64	0.09	0.12	0.03
PHY	0.88	0.10	12.13	4.45	2.42	0.44	0.74	0.09	0.14	0.04
PSY	0.82	0.11	13.26	4.61	2.53	0.45	0.66	0.09	0.14	0.04
ST	0.86	0.10	13.84	4.53	2.59	0.44	0.71	0.09	0.13	0.03
ALL	0.85	0.10	13.71	4.84	2.58	0.47	0.71	0.09	0.14	0.04

Appendix 4*. Lexical variation – lexical word variation and verb variation*

Indices	*NV*		*ADJV*		*ADVV*		*MODV*	
R.areas	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*
AA	0.64	0.10	0.17	0.04	0.04	0.02	0.21	0.05
BE	0.63	0.11	0.15	0.04	0.04	0.02	0.19	0.05
BIO	0.64	0.10	0.17	0.04	0.04	0.02	0.21	0.05
CHEM	0.68	0.11	0.17	0.05	0.04	0.02	0.22	0.05
CS	0.64	0.11	0.16	0.04	0.05	0.02	0.20	0.05
ECO	0.60	0.09	0.15	0.03	0.05	0.02	0.20	0.04
EDU	0.62	0.10	0.15	0.03	0.04	0.02	0.19	0.04
ENG	0.63	0.10	0.16	0.04	0.04	0.02	0.20	0.04
EOH	0.56	0.09	0.14	0.03	0.03	0.02	0.18	0.04
MS	0.67	0.10	0.17	0.04	0.04	0.02	0.21	0.05
MATH	0.67	0.12	0.18	0.05	0.04	0.03	0.22	0.05
NN	0.57	0.10	0.16	0.04	0.04	0.02	0.20	0.05
PHY	0.68	0.10	0.17	0.05	0.04	0.03	0.21	0.06
PSY	0.58	0.10	0.16	0.04	0.04	0.02	0.20	0.05
ST	0.64	0.11	0.17	0.04	0.05	0.02	0.22	0.05
ALL	0.64	0.11	0.16	0.04	0.04	0.02	0.21	0.05

Appendix 5. *Lexical variation – Noun, adjective, adverb, and modifier variation*

Indices	*MLS*		*MLT*		*MLC*	
R.areas	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*
AA	26.55	10.39	26.22	10.31	19.77	9.35
BE	24.31	5.49	23.35	5.74	15.59	4.29
BIO	24.68	6.23	23.79	6.84	17.18	4.63
CHEM	25.70	8.26	24.37	7.96	18.65	6.39
CS	23.73	10.35	22.83	10.46	16.78	6.84
ECO	24.99	5.80	23.56	5.68	16.99	4.16
EDU	23.53	5.68	22.51	5.85	15.36	3.57
ENG	23.30	5.32	22.27	5.45	17.13	4.49
EOH	23.93	6.42	24.88	8.11	17.70	4.34
MS	24.48	9.06	23.45	9.19	17.93	8.64
MATH	26.90	21.97	25.72	21.89	18.34	18.43
NN	23.60	7.59	24.08	8.59	17.52	5.69
PHY	26.09	10.28	25.05	10.24	19.37	9.82
PSY	24.23	6.88	23.24	6.90	16.24	5.75
ST	23.97	4.69	22.74	4.68	17.21	4.09
ALL	24.67	8.41	23.72	8.52	17.56	7.12

Appendix 6. *Length of production unit*

Indices	*C/S*	
R.areas	*mean*	*sd*
AA	1.41	0.40
BE	1.62	0.39
BIO	1.48	0.32
CHEM	1.42	0.34
CS	1.46	0.37
ECO	1.51	0.31
EDU	1.57	0.39
ENG	1.40	0.29
EOH	1.39	0.33
MS	1.41	0.31
MATH	1.56	0.50
NN	1.38	0.30
PHY	1.41	0.37
PSY	1.56	0.40
ST	1.44	0.32
ALL	1.46	0.35

***Appendix** 7. Complex structures*

Indices	*C/T*		*CT/T*		*DC/C*		*DC/T*	
R.areas	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*
AA	1.38	0.34	0.33	0.21	0.27	0.15	0.42	0.30
BE	1.55	0.38	0.42	0.21	0.34	0.14	0.57	0.35
BIO	1.41	0.31	0.33	0.20	0.26	0.13	0.40	0.28
CHEM	1.34	0.30	0.28	0.20	0.23	0.14	0.34	0.28
CS	1.40	0.34	0.34	0.20	0.28	0.14	0.43	0.31
ECO	1.41	0.28	0.34	0.18	0.28	0.12	0.42	0.26
EDU	1.49	0.34	0.38	0.21	0.31	0.13	0.49	0.31
ENG	1.33	0.26	0.29	0.18	0.24	0.13	0.34	0.24
EOH	1.42	0.32	0.23	0.18	0.22	0.13	0.34	0.26
MS	1.34	0.27	0.29	0.19	0.24	0.13	0.35	0.25
MATH	1.48	0.45	0.39	0.24	0.30	0.17	0.50	0.38
NN	1.39	0.29	0.25	0.18	0.23	0.13	0.35	0.25
PHY	1.35	0.32	0.30	0.21	0.24	0.15	0.37	0.30
PSY	1.48	0.33	0.36	0.21	0.30	0.14	0.48	0.32
ST	1.35	0.27	0.31	0.19	0.25	0.13	0.37	0.26
ALL	1.39	0.32	0.32	0.20	0.26	0.14	0.40	0.29

***Appendix 8**. Subordination*

Indices	CP/C		CP/T		T/S	
R.areas	mean	sd	mean	sd	mean	sd
AA	0.45	0.25	0.59	0.30	1.02	0.14
BE	0.57	0.36	0.84	0.47	1.05	0.13
BIO	0.58	0.31	0.79	0.39	1.05	0.13
CHEM	0.58	0.34	0.75	0.40	1.06	0.13
CS	0.52	0.35	0.68	0.37	1.05	0.13
ECO	0.61	0.31	0.83	0.41	1.07	0.12
EDU	0.59	0.31	0.86	0.43	1.06	0.15
ENG	0.53	0.30	0.69	0.37	1.06	0.12
EOH	0.67	0.32	0.93	0.47	0.99	0.20
MS	0.54	0.32	0.69	0.37	1.06	0.12
MATH	0.35	0.33	0.47	0.35	1.06	0.16
NN	0.58	0.28	0.79	0.38	1.00	0.16
PHY	0.48	0.31	0.61	0.35	1.05	0.14
PSY	0.62	0.47	0.86	0.47	1.06	0.16
ST	0.59	0.33	0.76	0.38	1.06	0.11
ALL	0.55	0.33	0.74	0.40	1.05	0.14

Appendix 9. *Coordination*

Indices	CN/C		CN/T		VP/T	
R.areas	mean	sd	mean	sd	mean	sd
AA	2.78	0.99	3.73	0.99	1.91	0.47
BE	2.38	0.75	3.55	0.96	2.27	0.58
BIO	2.68	0.71	3.72	1.00	1.97	0.51
CHEM	2.82	0.81	3.69	0.98	1.85	0.44
CS	2.44	0.76	3.30	0.93	2.11	0.50
ECO	2.58	0.66	3.58	0.91	1.97	0.46
EDU	2.27	0.65	3.33	1.01	2.21	0.64
ENG	2.55	0.73	3.32	0.91	1.92	0.45
EOH	2.56	0.71	3.61	1.30	2.01	0.51
MS	2.76	0.79	3.64	1.02	1.86	0.43
MATH	2.45	0.85	3.47	1.18	2.05	0.63
NN	2.64	0.64	3.65	1.09	1.93	0.46
PHY	2.93	0.91	3.81	1.03	1.89	0.47
PSY	2.40	0.82	3.44	1.05	2.11	0.55
ST	2.72	0.75	3.60	0.92	1.92	0.43
ALL	2.64	0.78	3.58	1.00	1.97	0.50

Appendix 10. *Particular structures*

	AA	BE	BIO	CHEM	CS	ECO	EDU	ENG	EOH	MATH	MS	NN	PHY	PSY	ST
LD	-0.06	0.03	**0.17**	0.13	-0.08	**0.10**	0.01	-0.01	-0.02	**0.33**	**0.21**	0.09	**-0.08**	0.11	0.18
LS1	**-0.13**	-0.02	**-0.15**	-0.07	**-0.17**	0.01	0.11	-0.01	-0.06	-0.20	-0.01	-0.09	**-0.09**	-0.01	-0.05
LS2	**-0.17**	0.01	-0.07	-0.04	-0.13	0.02	0.03	-0.04	-0.08	-0.11	0.07	-0.04	**-0.11**	-0.03	-0.04
VS1	-0.06	0.05	0.06	0.00	-0.01	0.02	-0.03	0.02	-0.05	0.10	0.07	-0.02	**0.09**	-0.08	**0.08**
VS2	0.01	0.03	0.11	-0.01	0.01	0.04	-0.02	0.03	-0.04	**0.21**	0.06	-0.01	**0.07**	-0.08	0.07
CVS1	0.01	0.03	0.11	-0.01	0.01	0.04	-0.02	0.03	-0.04	**0.21**	0.06	-0.01	**0.07**	-0.08	0.07
NDW	0.34	0.00	0.03	-0.04	0.09	0.04	0.00	0.11	0.09	**0.23**	**0.12**	-0.03	0.10	-0.02	-0.06
NDWZ	0.07	-0.09	**0.14**	**0.14**	0.09	**0.08**	-0.05	0.10	0.05	0.10	0.24	0.04	**0.12**	-0.01	0.16
NDWERZ	**0.13**	-0.06	**0.16**	0.22	-0.06	0.17	-0.02	**0.18**	-0.01	0.10	0.28	**0.12**	**0.08**	-0.05	0.20
NDWESZ	0.09	**-0.16**	**0.15**	**0.18**	-0.01	0.19	-0.03	**0.16**	-0.07	0.16	**0.21**	**0.13**	0.06	0.03	0.22
TTR	**-0.15**	-0.03	0.11	**0.16**	-0.13	**0.09**	-0.07	0.09	-0.06	-0.07	**0.21**	0.07	0.02	-0.05	0.23
MSTTR	0.09	**-0.13**	0.16	0.20	-0.02	0.19	-0.09	**0.12**	-0.03	0.16	0.25	0.09	**0.09**	0.00	0.24
CTTR	0.32	-0.02	0.11	0.05	0.62	**0.09**	-0.03	**0.17**	0.05	**0.24**	0.25	0.03	**0.13**	-0.04	**0.11**
RTTR	0.32	-0.02	0.11	0.05	0.05	**0.09**	-0.03	**0.17**	0.05	**0.25**	0.25	0.03	**0.13**	-0.04	**0.11**
LOGTTR	-0.07	-0.02	**0.12**	**0.16**	-0.10	**0.11**	-0.09	**0.13**	-0.02	-0.05	0.24	0.07	0.04	-0.06	0.24
UBER	0.05	-0.03	**0.14**	**0.18**	-0.07	0.13	-0.06	**0.17**	0.01	0.04	0.28	0.07	**0.08**	-0.05	0.23
LV	-0.08	-0.01	-0.05	0.09	-0.12	-0.03	-0.02	0.02	-0.05	-0.10	**0.14**	-0.04	0.04	**-0.13**	**0.08**
VV1	0.22	-0.01	0.16	0.06	0.03	0.04	0.05	0.08	-0.04	**0.30**	0.08	-0.02	0.06	-0.09	0.02
SVV1	0.22	-0.01	0.16	0.06	0.03	0.04	0.05	0.08	-0.04	**0.30**	0.08	-0.01	0.06	-0.09	0.02
CVV1	**-0.12**	-0.04	0.09	**0.12**	-0.12	0.02	-0.09	0.06	-0.10	-0.13	**0.20**	0.02	0.03	-0.11	0.14
VV2	**-0.12**	-0.05	0.24	**0.11**	-0.01	0.06	0.03	0.05	-0.12	0.09	-0.05	-0.03	0.01	-0.04	**0.10**
NV	**-0.12**	-0.04	0.08	**0.10**	**-0.15**	-0.01	-0.06	0.08	-0.07	-0.16	**0.17**	0.03	0.03	-0.09	**0.10**
ADJV	0.03	0.01	0.08	**0.14**	-0.04	0.17	-0.09	0.01	0.03	0.03	**0.18**	0.01	-0.01	-0.05	0.22
ADVV	**0.11**	0.09	-0.01	**0.14**	0.10	0.06	0.00	**0.13**	-0.05	0.05	**0.11**	-0.06	-0.02	0.06	**0.11**
MODV	0.08	0.06	0.09	0.20	-0.01	0.16	-0.06	0.09	-0.04	0.02	**0.20**	0.00	-0.01	-0.03	0.25
N	336	288	302	457	176	731	173	301	217	91	313	392	773	274	778

Appendix 11. *Correlation lexical complexity and Altmetric Attention Score. Statistically significant at p ≤ 0.05 (bold)*

	AA	BE	BIO	CHEM	CS	ECO	EDU	ENG	EOH	MATH	MS	NN	PHY	PSY	ST
MLS	-0.02	-0.05	0.03	-0.02	0.67	0.04	**0.16**	-0.06	0.09	-0.10	-0.04	0.03	0.04	-0.05	0.07
MLT	0.01	-0.04	0.00	-0.01	0.65	0.03	**0.21**	-0.04	0.06	0.05	-0.01	-0.06	0.04	-0.07	0.03
MLC	-0.05	0.09	-0.05	-0.01	0.62	-0.02	0.07	-0.03	0.11	0.17	0.04	0.03	0.07	-0.04	-0.06
C/S	0.00	**-0.12**	0.04	-0.01	0.69	0.06	0.08	0.02	0.03	**-0.25**	-0.06	-0.01	-0.04	-0.02	**0.11**
VP/T	-0.04	-0.07	**0.12**	0.06	0.67	**0.08**	0.08	0.02	-0.05	0.16	-0.04	-0.09	**-0.11**	-0.01	**0.08**
C/T	0.04	**-0.13**	0.02	0.02	0.68	0.06	0.14	0.04	-0.05	-0.13	-0.06	**-0.16**	-0.05	-0.01	**0.11**
DC/C	0.03	-0.11	0.05	0.06	0.64	0.10	0.09	0.03	0.03	-0.16	-0.05	-0.04	**-0.09**	0.04	0.17
DC/T	0.04	**-0.12**	0.04	0.05	0.65	**0.09**	0.11	0.03	0.01	-0.16	-0.06	-0.07	**-0.08**	0.03	0.15
T/S	-0.06	0.05	0.01	-0.05	0.68	-0.01	0.00	-0.02	0.11	**-0.26**	-0.04	**0.13**	0.01	0.04	0.05
CT/T	0.02	-0.12	0.05	0.06	0.65	0.07	0.06	0.04	0.06	-0.16	0.00	-0.02	-0.07	0.00	0.17
CP/T	0.08	-0.02	0.04	-0.05	0.67	-0.01	0.07	0.03	-0.01	**0.21**	0.10	0.06	0.04	0.07	0.03
CP/C	0.03	0.02	0.02	-0.04	0.66	-0.03	0.02	0.01	0.01	**0.24**	**0.12**	**0.10**	0.06	0.07	0.00
CN/T	-0.04	-0.07	0.04	0.04	0.58	0.06	**0.16**	-0.04	0.05	0.05	0.01	-0.08	0.00	-0.03	**0.09**
CN/C	**-0.13**	0.02	-0.01	0.03	0.57	0.00	0.08	-0.02	0.12	0.18	0.04	0.01	0.02	-0.04	0.02
N	336	288	302	457	176	731	173	301	217	91	313	392	773	274	778

Appendix 12. *Correlation syntactic complexity and Altmetric Attention Score. Statistically significant at p ≤ 0.05 (bold)*

	AA	BE	BIO	CHEM	CS	ECO	EDU	ENG	EOH	MATH	MS	NN	PHY	PSY	ST
LD	0.248	0.595	0.003	0.006	0.272	0.007	0.851	0.872	0.809	0.001	0.000	0.078	0.029	0.073	0.000
LS1	0.015	0.743	0.008	0.139	0.023	0.775	0.145	0.864	0.389	0.053	0.869	0.077	0.014	0.818	0.184
LS2	0.002	0.810	0.257	0.366	0.079	0.666	0.650	0.506	0.269	0.295	0.247	0.406	0.003	0.589	0.255
VS1	0.293	0.440	0.301	0.931	0.872	0.680	0.673	0.709	0.428	0.334	0.189	0.660	0.013	0.166	0.027
VS2	0.875	0.629	0.054	0.756	0.908	0.316	0.772	0.583	0.518	0.046	0.261	0.818	0.044	0.197	0.051
CVS1	0.878	0.616	0.055	0.765	0.910	0.318	0.763	0.563	0.527	0.048	0.258	0.818	0.045	0.199	0.051
NDW	0.000	0.992	0.549	0.383	0.214	0.320	0.958	0.050	0.195	0.028	0.028	0.594	0.005	0.754	0.111
NDWZ	0.183	0.148	0.019	0.002	0.236	0.023	0.549	0.090	0.481	0.369	0.000	0.412	0.001	0.897	0.000
NDWERZ	0.021	0.289	0.005	0.000	0.413	0.000	0.790	0.002	0.903	0.338	0.000	0.016	0.024	0.429	0.000
NDWESZ	0.094	0.006	0.009	0.000	0.910	0.000	0.680	0.006	0.335	0.137	0.000	0.013	0.112	0.598	0.000
TTR	0.008	0.575	0.059	0.000	0.093	0.012	0.347	0.122	0.388	0.518	0.000	0.181	0.571	0.426	0.000
MSTTR	0.086	0.028	0.005	0.000	0.828	0.000	0.248	0.040	0.706	0.135	0.000	0.072	0.015	0.978	0.000
CTTR	0.000	0.718	0.055	0.247	0.000	0.013	0.701	0.003	0.459	0.019	0.000	0.522	0.000	0.543	0.003
RTTR	0.000	0.720	0.056	0.250	0.532	0.013	0.706	0.003	0.460	0.019	0.000	0.527	0.000	0.540	0.003
LOGTTR	0.202	0.763	0.033	0.001	0.203	0.002	0.226	0.030	0.730	0.650	0.000	0.181	0.246	0.317	0.000
UBER	0.319	0.578	0.015	0.000	0.363	0.001	0.446	0.003	0.846	0.710	0.000	0.148	0.032	0.384	0.000
LV	0.159	0.880	0.424	0.069	0.111	0.460	0.825	0.704	0.465	0.350	0.012	0.484	0.252	0.027	0.031
VV1	0.000	0.904	0.005	0.235	0.712	0.237	0.512	0.180	0.537	0.003	0.177	0.765	0.091	0.121	0.522
SVV1	0.000	0.904	0.006	0.225	0.702	0.241	0.523	0.181	0.546	0.004	0.177	0.768	0.084	0.119	0.513
CVV1	0.035	0.485	0.120	0.011	0.108	0.527	0.247	0.305	0.144	0.233	0.000	0.727	0.349	0.074	0.000
VV2	0.027	0.422	0.000	0.024	0.850	0.104	0.695	0.431	0.079	0.376	0.334	0.607	0.752	0.471	0.006
NV	0.024	0.469	0.182	0.031	0.046	0.708	0.419	0.182	0.290	0.121	0.003	0.523	0.470	0.156	0.004
ADJV	0.641	0.930	0.141	0.003	0.566	0.000	0.224	0.851	0.674	0.803	0.002	0.833	0.871	0.387	0.000
ADVV	0.038	0.135	0.860	0.003	0.192	0.086	0.981	0.019	0.428	0.637	0.044	0.241	0.510	0.316	0.002
MODV	0.151	0.328	0.117	0.000	0.926	0.000	0.402	0.102	0.574	0.829	0.000	0.959	0.801	0.670	0.000

Appendix 13. *P-values for lexical complexity*

	AA	BE	BIO	CHEM	CS	ECO	EDU	ENG	EOH	MATH	MS	NN	PHY	PSY	ST
MLS	0.657	0.434	0.626	0.700	1.994	0.325	0.032	0.271	0.164	0.347	0.481	0.586	0.234	0.384	0.067
MLT	0.880	0.546	0.948	0.909	4.289	0.355	0.006	0.459	0.412	0.663	0.891	0.252	0.279	0.245	0.342
MLC	0.344	0.132	0.353	0.750	2.575	0.642	0.395	0.645	0.107	0.110	0.531	0.501	0.066	0.522	0.108
C/S	0.949	0.040	0.544	0.837	3.746	0.105	0.277	0.705	0.708	0.018	0.254	0.882	0.236	0.711	0.001
VP/T	0.489	0.229	0.030	0.219	2.335	0.022	0.308	0.754	0.433	0.127	0.469	0.074	0.002	0.858	0.019
C/T	0.459	0.030	0.687	0.665	1.279	0.086	0.059	0.471	0.451	0.233	0.294	0.002	0.179	0.894	0.003
DC/C	0.640	0.075	0.375	0.229	5.551	0.008	0.228	0.630	0.643	0.121	0.378	0.410	0.015	0.551	0.000
DC/T	0.497	0.044	0.505	0.302	3.513	0.015	0.146	0.575	0.896	0.127	0.309	0.186	0.027	0.635	0.000
T/S	0.297	0.420	0.859	0.248	3.933	0.794	0.951	0.793	0.099	0.014	0.501	0.009	0.820	0.461	0.204
CT/T	0.677	0.051	0.356	0.198	3.107	0.054	0.410	0.468	0.394	0.141	0.993	0.681	0.061	0.988	0.000
CP/T	0.168	0.726	0.512	0.299	6.153	0.842	0.354	0.639	0.889	0.049	0.077	0.265	0.273	0.218	0.471
CP/C	0.549	0.673	0.711	0.396	3.203	0.395	0.806	0.922	0.872	0.019	0.042	0.038	0.101	0.246	0.954
CN/T	0.484	0.244	0.530	0.423	2.18	0.096	0.030	0.456	0.483	0.629	0.794	0.109	0.952	0.601	0.011
CN/C	0.020	0.791	0.853	0.576	1.144	0.903	0.281	0.715	0.083	0.093	0.493	0.835	0.531	0.530	0.636

Appendix 14. *P-values for syntactic complexity*