Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The Dynamics of Collaboration: A Scientometric Study of Team Size and Text Complexity in Scientific Abstracts from Bangladesh

Ravikumar S.
Central University of Tamil Nadu
Mezbah-ul-Islam Muhammad
University of Dhaka
Chandrappa ,
chandruekm@gmail.com

Indian Maritime University

Research Article

Keywords: team size, scientific collaboration, readability, lexical diversity, text analysis, scientometrics, collaborative writing, research team dynamics, Bangladesh

Posted Date: March 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8251341/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

The geography of scientific collaboration has become increasingly internationalized, transforming where and how research is constructed and communicated. While collaboration has been shown to impact scientific production and impact, much less is known about how it interacts with the linguistic characteristics of scientific communication. This study analyses the extent to which team size has an impact on the linguistic features of scientific abstracts from scholars with Bangladesh affiliations. Using the full population of 25,883 WoS publications (1972–2023), I assess the differences in three relevant linguistic characteristics - readability, lexical diversity, and abstract length - based on collaboration type (solo, small groups (2 authors), medium groups (3–5), large groups (6–10) and very-large groups (> 10 authors)). Readability is assessed through the Flesch Reading Ease, Flesch-Kincaid Grade Level and Gunning FOG Index while linguistic diversity is assessed through the Type-Token Ratio (TTR). The findings reveal three primary conclusions. First, lexical diversity is negatively impacted by team size in which greater linguistic inclusivity was found within collaborative endeavour's as team size increased. Second, readability behaves according to a curvilinear pattern whereby 6–10 author teams produce the easiest-to-read abstracts and > 10 author teams produce more concise, complex works. Third, abstract length is positively impacted by team size due to the broader conceptual and methodological scope of collaborative efforts. Mean differences for readability relative to team size are statistically significant (ANOVA; $p < 0.001$) as a negative trend of reading ease from 2019–2024 reveals a decrease for all group sizes.

This study contributes to the emerging field of Scientometrics - by connecting structures of collaboration to assessable linguistic features in practical applied scientific output for communication purposes. Practical applications are made for research evaluation, journal editors and writing specialists who teach novice researchers in burgeoning research fields how to produce such collaborative findings effectively while best anticipating expectations. Ultimately, these results indicate that researchers should welcome access to linguistic features in evaluation of collaborative efforts as communication does not come without cost when collaborative scientific writing occurs.

Introduction

Scientific collaboration is an increasingly central component of twenty-first century knowledge construction. Cross-disciplinary researchers work alongside one another to engage at a level of complementary expertise, access to resources, and methodological ability that is necessary for increasingly complicated scientific inquiries. Supported by increasingly more international opportunities for publication and research output, findings increasingly boast publication trends, co-authorship tendencies, increased team size and extensive networks to paper publication. Thus, thus far, collaboration has come to influence productivity, and consequently, knowledge transmission; however, collaboration increasingly connects with how scientific knowledge comes to be written, presented, and subsequently disseminated.

At the same time, an increased scholarly focus exists relative to the language characteristics of scientific communication. Recent interdisciplinary findings note that the readability of scientific papers has increased over time relative to lexicon, organization of text, discipline specific jargon and standards of scientific communications established over twenty-first century developments. Text complexity - readability, lexical diversity and structural complexity - provides insight into how discoverable works of scientific communications relative to their complexities are or are not accessible to average audiences for comprehension, citation and application purposes. Yet despite such growing interest across academia and publication, the role of collaboration in this process remains relatively unexplored. Many studies assess team science relative to productivity, impact considerations. Still, few consider how team size (the level or absence of collaboration on authorship) - the number of authors - can impact conciseness of argument, depth of diction and informational value across science-related abstracts.

Bangladesh represents a timely and relevant space for investigating this further. As a developing country, Bangladesh boasts rapid growth relative to its system of research within the last two decades - growing publication output, growing levels of in-country co-authorship, and growing interdisciplinary connections characterize the developmental and academic/publishing realities of researchers and students across disciplines. Public and private universities support this output while public/government laboratories and specialty research institutes support interdisciplinary connections in the biomedical sciences, engineering, agriculture, public health, and social sciences. However, thus far, no study has ventured to connect how the shifting nature of collaboration within Bangladesh has shifted over time to impact authorial writing and literacy characteristics in the specific features of scientifically published writing.

Abstracts represent a potent space for assessing such change; as relatively shorter pieces to longer works representing content-based papers, abstracts wield methodological appeal via necessity and relevant framing which subsequently inform their own structure and complexity. Abstracts can be simple or complex depending on where they are sent or to what audience they are geared linguistically. Thus, word choice, length and ensuing complexity can come from a collaborative effort of team dynamics which leverage negotiation amongst diverse voices or from limited dynamics.

Thus far, research exists relative to scientific impact and team size (Larivière et al., 2015); for example, while those teams of larger size accrue the most citations over time, the smaller sized teams present novel concepts or disruptive findings (Wu, Wang & Evans, 2019). Comparative studies note readability - Plavén-Sigray et al., (2017) - is decreasing over time generally. Yet little effort has been exerted to assess how these two advancements are connected; does team size impact readability? Does the team size on a spectrum operate linearly when it comes to language features?

This project aims to fill that void between established contributions to confirm or deny observations related to collaborative writing across developed systems where they join established literature in scientometrics relative to effective team science practices and more generally applicable contributions relevant to scholarly communication. A large corpus study is employed with 25,883 abstracts indexed by Web of Science from 1972–2023 with authors from institutions within Bangladesh which explores

length, readability and lexical diversity across different collaboration size tiers, specifically one author. One author collaboration exists.

Review of Literature

Collaboration Dynamics and Global Scientific Output

Collaborative science has shifted considerably over decades in response to levels of specialization, greater facility ability, and international competition. The team-oriented nature of research functions as the primary means of knowledge production for any given field and channels the configuration of outputs, patterns of impact and communication. Internationally, large-scale studies indicate an increasingly progressive, global trend toward expanded collaboration. Gazni, Sugimoto and Didegah (2011) cite significant increases in small and medium collaborative teams in a sample of over 14 million records in the Web of Science while Adams et al. (2005) find a 50% increase in collaborative team size in U.S. universities from 1981–1999. Such findings help establish the notion that science exists in a highly complexly structured, interdisciplinary setting today where institutions facilitate this collaborative means of research.

This is especially true in emerging research countries like Bangladesh where collaboration patterns, institutional capacity and age of research and communication patterns develop in a new, rapidly growing setting.

Growth of Team Science and Evolution of Collaboration Structures

Such cumulative findings support unit size and collaborative component due to technological, institutional, national and disciplinary factors over the past twenty years. For example, Gazni et al. (2011) note that various increases were noted in small- and medium-size groups as technological advancements allow information to be sent to larger ranges without major costs reducing for necessary or Adams et al. (2005) suggests that although collaboration is high (even within the U.S.), levels which boast significant increased over the course of the twenty-year study boast highly productive increases of papers within the U.S.

According to Gazni and Didegah (2011), Harvard University boasts 88% of papers produced with collaborative efforts and papers produced across institutions boast extremely high levels of citations as well. Furthermore, Franceschet and Costantini (2010) relative to 20 different disciplines studied in Italy reveal that collaboration, team size, potential for citation and potential for peer-review increase in a positive fashion, as do all variables.

Yet there's a variable - discipline. For example, Abt (2007) asserts that international collaboration via publication is field dependent - international collaboration accounted for 13% of published papers in surgery in 2007 while 55% of papers in astronomy - pointing to some geopolitical interdisciplinary

epistemology at play and Didegah & Gazni (2014) assess national collaboration within niche fields while international collaboration is an asset for international visibility and eventual citation potential.

Overall, these cumulative findings suggest that unit size and collaborative component is based on technological, institutional, national and disciplinary factors - all pertinent to the specific research field in Bangladesh's developing research.

Collaboration Networks and Their Influence on Impact and Innovation

But network structure is just as important as network size. Liu et al. (2024) note that the more "small-world characteristics" (high clustering and low path length) teams have, the more they influence publications. Furthermore, levels of national collaboration reliance vary. For instance, the Scientific Dependence Index shows that the smaller systems (Cuba, for instance) rely more heavily upon international collaboration while larger systems boast more balanced output - although this, too, is subject to geopolitical and historical developments.

Likewise, in Europe, Zitt et al. (2020) found that long-term political, linguistic and historical connections help understand how collaboration occurs on a bilateral basis. Thus, the fact that Bangladesh is trending more reliant on international networks over time suggests levels of potentially impacted research novelty and language use in scientific publications.

But collaboration isn't always helpful. Mongeon and Larivière (2015) found that collaborators of researchers responsible for retracted publications experience drastic declines in future publication rates and citation impact which indicates caution when deciding to co-author with someone else.

Ultimately, while many authors today contribute to discoveries, small author teams have breakthroughs uniquely tied to them. Winnink et al. (2015) found via the development of the ubiquitin-mediated proteolytic system that concentrated collaboration results in accelerated conceptizations over time.

Text Complexity, Readability, and Scientific Communication

Research on linguistic characteristics of scientific literature contributes to an additional layer of exploring scientific communication. According to Dressler et al. (2025), readability is relative, based upon the alignment of linguistic characteristics with textual density and cognitive burden. The denser the language used, the more cognitive effort required, and the lower the potential for comprehension.

Findings relative to scientific writing implicate decline in readability, regardless of discipline. For example, in the findings relative to plain language summaries, it was determined by Lang et al. (2025) that many plain language summaries fell above recommended levels of readability and discipline-specific terms frequently connected to interdisciplinary efforts and large research teams.

Plavén-Sigray et al. (2017) report a similar decline in scientific readability from 1881 to 2015 across 12 dominant fields of inquiry. Furthermore, linguistic findings support that collaborative writing reduces lexical variety and encourages convergence into homogeneity of word choice (Lu et al., 2019; Song et al., 2023).

Therefore, this data supports that collaborative writing may have an impact on the density of linguistic characteristics.

Research Context and Gaps in Bangladesh

However, with Bangladeshi research on the rise, the following is true for the literature:

1. Many bibliometric works have not accounted for collaborative information across institutions of Bangladesh.
2. No corroborating empirical research exists for the reader effect of group size and text complexity (readability, lexical diversity).
3. No scientifically standardized national scope scientometric study exists that links collaboration and linguistic variables on a sample of scientific abstracts.
4. Few studies exist that assess the ability of authors to shift their writing strategies in accordance with collaborative means.

Thus, as research comes out of Bangladesh at a faster pace - higher collaborative rates, more international collaborations and institutions of varying strengths - now is the time to explore the nuances of abstract construction.

Synthesis and Conceptual Framework

These three themes emerge from the above literature review:

1. Team size and collaborative configuration affect scientific visibility and citations and impact measures, novelty (Franceschet & Costantini, 2010; Gazni & Didegah, 2011; Liu et al., 2024).
2. Collaboration patterns suggest research capacity exists based on previous engagement with certain countries and field tendencies (Abt, 2007; Didegah & Gazni, 2014; Ronda-Pupo, 2025).
3. Lexical complexity suggests cognitive exertion and communicative success, which comes from team size and collaborative writing effort (Dressler et al., 2025; Lang et al., 2025).

Therefore, these three themes of literature support the idea that team size will correlate with abstract clarity, lexical diversity and length, especially within an underdeveloped research context such as Bangladesh.

Collaboration Structures and Citation Impact

The findings of the research are overwhelmingly positive. Externally, research from external sources (Gazni & Didegah, 2011) and research from select universities across Italy (Franceschet & Costantini, 2010) indicates that where collaboration is concerned - from group size to citation usefulness - view it as a favorable, correlational study. Internally, the field research (Abt, 2007; Xu et al., 2021) indicates that increased collaboration in the personal lives of many people only improves levels of collaboration in research.

However, where collaboration and impact are concerned, little research exists in comparison to team science and linguistic or written pertinence to one paper. For example, impact relative to abstract familiarity or lexical diversity has not been researched as it pertains to Bangladesh for no relative research exists to date.

Therefore, this research is necessary.

Objectives of the Study

This study aims to:

1. Investigate collaborative tendencies in the linguistic construction of Bangladeshi scientific abstracts.
2. Examine levels of linguistic complexity (readability, lexical variety, syntactic complexity and semantic complexity) in Bangladeshi scientific abstracts.
3. Establish whether team size is an influencing factor for shifts in textual complexity.
4. Discuss socio-academic drivers of collaboration within Bangladesh.

Methods

Data Source and Pre-processing

The dataset used for the study was created from the WoS (Web of Science) Core Collection. It encompasses all publications that have at least one author affiliated with a Bangladeshi institution. The first step of data cleaning was performed in R (v4.3.1). Records without abstracts, empty abstracts and abstracts below 50 characters in length were removed from the study - this ensures that the papers are sufficient for a large enough language study. The study's final analytical dataset consisted of 25,883 scientific abstracts after this stage.

The study estimated team size by assessing the number of authors present in the WoS field designated for authors. For example, authors separated by semicolon, semi-colon, were each counted for contribution; those without any data were counted as one author. Thus, the publications were classified into a type of collaboration category: *Solo* (1 author), *Small team* (2 authors), *Medium team* (3–5

authors), *Large team* (6–10 authors), *Very large team* (> 10 authors). These categories were treated as ordered factors in further analysis.

Linguistic Feature Extraction

Languaging was conducted via the quanteda package (Benoit et al., 2018). A corpus object was created from the authenticated abstracts with each abstract being its own document.

Length and Structural Metrics

The following textual variables were calculated per document.

- Total number of words: total number of character strings excluding whitespace
- Total number of sentences: estimated via terminal punctuation (., !, ?)

They were standards of equivalence approximations for extent and density of form.

Readability Measures

Readability was assessed using textstat_readability(). Three widely used indices were computed

1. Flesch Reading Ease (FRE): scores from 0–100, where higher scores indicate greater readability.
2. Flesch–Kincaid Grade Level (FKGL): estimates the U.S. school grade required to comprehend the text.
3. Gunning FOG Index: estimates years of formal education needed for first-pass understanding.

Lexical Diversity

The Type-Token Ratio (TTR) was calculated to represent lexical diversity. The same tokenization process occurred where the abstracts' punctuation, numbers, and special characters were removed, and the case was transformed into lowercase. Then, textstat_lexdiv() was applied to measure this variable

$$\mathrm{TTR} = \frac{\text{Unique word types}}{\text{Total word tokens}}$$

Statistical Analysis and Visualization

All tests were run in R. Descriptive statistics (mean, sd) for all linguistic variables were acquired across all five team-size groups using the dplyr package.

Correlation Matrix

To assess linear relationships between team size and abstract length (tokens) versus readability measures and lexical diversity, a Pearson correlation matrix was constructed.

Group Comparisons

A one-way ANOVA was performed to compare the means of linguistic features across collaboration groups (where FRE was the dependent variable, and team size category was the grouping factor). This assessed whether different levels of collaboration impact readability.

Visualizations

All visualizations were constructed with ggplot2, which subsequently had a publication-ready theme applied. ggplot2 created the following output

- A bar chart visualizing team sizes of papers with frequency
- A violin + box plot showing the distributions of the readability/length and lexical variables
- A heat map denoting correlations between all linguistic variables
- Line plots per year (2019–2024)

Visualizations helped interpret any collaboration-linguistic relationships.

Results

Team Size Distribution

An investigation into the 25,883 works reveals a shift towards more collaborative authorship. According to Fig. 1, medium (3–5 authors) and large (6–10 authors) collaborative efforts dominate as 10,165 (39%) and 9,869 (38%) papers were published by medium and large collaborative efforts, respectively. Only 500 of the 25,883 works (approximately 2%) represent solo efforts.

This reflects, to a degree, the collaborative efforts of publication from Bangladesh as over time, more and more works are published by teams rather than individuals.

Team Size and Abstract Characteristics

Interestingly, a specific relationship emerged between team size and abstract structure. Not only did grand means of an abstract length greater for team size, but the means of word counts per level of team size occurred more frequently, too. The grand mean was 198.62 words for 1 author and 266.99 for extremely large teams (> 10). This means that the larger the team, the more extensive and more likely and encompassing abstract sent is as more people contribute more information to include.

In contrast, lexical diversity decreased by team size on the basis of Type-Token Ratio (TTR). The average TTR score of the smaller (1.54), (1) and the extremely large teams (> 10) (0.56). This means that there are less lexically diverse options made in collaboration abstractions than there are in those made by one author.

Table 1
Summary of Abstract Metrics by Team Size Category

Team Size Category	N Papers	Avg. Word Count	Avg. Flesch Score	Avg. F-K Grade	Avg. TTR
Solo (1)	500	198.62	19.98	16.74	0.59
Small team (2)	1773	206.28	21.02	16.47	0.58
Medium team (3–5)	10165	218.91	21.1	16.51	0.58
Large team (6–10)	9869	237.03	21.58	16.52	0.57
Very large team (> 10)	3576	266.99	20.45	16.82	0.56
Note. F-K Grade = Flesch-Kincaid Grade Level; TTR = Type-Token Ratio.					
Two trends are particularly evident:					

- Abstract length increases with team size (Δ ≈ +68 words between solo teams and very large teams)
- Lexical diversity decreases as TTR scores decline.

These two conclusions indicate that the greater the collaborative effort, the greater the likelihood of a longer - but lexically unimpressive - abstract.

Team Size and Readability

Readability was related to team size, non-linearly. A one-way ANOVA established differences in FRE across the five types of collaboration, $F(4, 25{,}878) = 13.55$, $p < .001$.

Ultimately, the most readable abstracts came from large collaborations (6–10 authors) with a mean FRE of 21.58. However, it's essential to realize that FRE decreases slightly for very large collaborations (> 10 authors) meaning that this trend approaches linguistic complexity at excessively collaborative stages.

The violin plots below in Fig. 3 show how FREs lie within each team, revealing much spread within groups but consistently low values in the absolute sense - typical of scientific prose where anything lower than 30 is considered technical, graduate student language.

Correlations between Linguistic Features

Correlation tests provided additional statistically significant observations between team size, word count, readability, and lexical diversity (see Fig. 4). For example, a moderately strong positive correlation existed between team size and word count ($r = .42$), indicating that the larger the team, the more words were used to construct the abstract. In addition, as anticipated, a slightly negative correlation existed between team size and lexical diversity (TTR) ($r = -.18$), since the more authorship dispersed in works,

the smaller the vocabulary used, a phenomenon observed in collaborative works and lexical convergence in scientometric research (Mongeon & Larivière, 2015; Winnink et al. 2015).

Metrics for readability told a different story. A correlation matrix found a correlation between FRE and word count (r = .21) and a correlation with team size (r = .14). This FRE finding in other words means that perhaps slightly longer papers create clarity of exposition. Furthermore, FRE also implies that teams larger than ten are conducive to a better understanding of average reading level materials with more authors. Similarly, FKGL and Gunning FOG Index are positively correlated with team size (r = .14 and r = .17 respectively). Thus, the more authors on a team the more complicated the writing is and required higher levels of reading comprehension. TTR is lexical diversity and this was negatively correlated with both the FOG Index (r = − .22) and FKGL (r = − .19) meaning papers with less lexically diverse vocabularies are also more complicated and conceptually dense. This makes sense in linguistic terms and with cognitive load theory (Mongeon & Larivière, 2015). Ultimately, these findings suggest a nuanced complex of collaborative and scientific writing.

Temporal Trends in Readability (2019–2024)

In order to measure whether the linguistic variables changed over time, a temporal assessment of the trend was performed for 2019–2024. The trend of FRE and TTR per team size is in Fig. 5.

The major findings from this assessment are as follows:

Decrease in Readability

FRE had an overall decrease over the five-year span (average ΔFRE = -1.8). This means increasingly dense scientific abstracts over the course of time years are less readable. This not only complements findings abroad that show decreases over time (Mongeon & Larivière, 2015; Winnink et al. 2015), but suggests that researchers are subject to these pressures internationally.

- *Decrease in Lexical Variety*: TTR showed a nominal decrease over this time frame (ΔTTR ≈ -0.02), meaning there is a standardization of vocabulary that likely means a homogeneity of writing style as well.
- *Team Size Distribution Consistency*: Regardless of annual increases and decreases in total published articles, the percent makeup of single author, small, medium, large and very large teams remained relatively proportionate and consistent with other years.

The increase of technicality and specialization to an increasingly less readable extent resonates with researchers internationally as international findings show a similar level of formality in interdisciplinary and increasingly specialized scientific communications. Therefore, the temporal findings suggest researchers similarly find themselves pressured over time by increasingly substantive events.

Discussion

This interdisciplinary study offers a comprehensive scientometric-linguistic analysis of the collaborative relationships that shape the linguistic characteristics of scientific abstracts from Bangladeshi authors. Drawing on the results of a dataset of 25,883 WoS publications, this study finds major characteristics where team size is appropriately matched to standardized type/token ratio (TTR), readability and length of abstract.

First, the relationship between team size and length of abstract is positively correlated. This implies that where authors are collaborating with more other authors, their abstracts become increasingly succinct presentations. Such an increase would be due to the ability for these collaborative contributions to find a balance between diverse approaches and larger designs necessitating increased understanding. Thus, with a multi-authored approach, a greater number of ideas can be posited, meaning that abstract length mirrors the complexity and intellectual comprehension expected of such analyses.

However, as team size increases, TTR decreases. This suggests that where authors are working together in a multi-author context, there is a similar lexicon used. This may be due to a consensus effect that teams who collaborated arrived at less specialized language that would be accessible in efforts to appeal to the diverse writing styles and discipline approaches. The results are aligned with findings of collaborative writing where groups of authors writing from drafts have less variation in lexical choices as authors work through a common voice. While this voice may be clearer, it is also homogenized to such an extent that it foregoes stylistic differences for eloquent economy that is more pronounced in larger teams.

The relationship between team size and readability is not necessarily linear. Teams of 6–10 authors have the highest readability - ironically compared to the extensive abilities of team dynamics - where it seems a happy median is found between enough persons for internal peer review and expanded editorial management yet not so many that there's a breakup in how to convey information. Excessive teams (> 10 authors) present with lower readability. This indicates that collaborative writing was not managed as effectively along with continued variations in writing and shifts in abstract purpose. This aligns with theoretical explorations of collaborative communication that suggests the larger the team, the less clear it becomes.

Second, the negative correlation to readability relative to time from 2019–2024 aligns with global trends. Researchers like Plavén-Sigray et al. (2017) found that from global access trends, general scientific writing has become less readable relative to increased specialization, author competitiveness and advanced focus on jargon use within the discipline. That Bangladesh finds itself amidst similar trends, however, suggests that as applied linguistic pressures penetrate modern science, technical density and complex methodologies and international publication patterns have permeated developed research fields just as much.

Ultimately, clear trends emerge whereby collaboration affects scientific writing in various directions. While larger teams do create longer abstracts, they also rely on standardized vocabulary; medium-sized teams create clearer works relative to small sized or excessively sized teams - while global literacy finds

itself increasingly downward trending in interdisciplinary endeavours suggests that the collaborative approaches are unlikely to be clear to greater public audiences.

These results are important for Bangladeshi scholars - a burgeoning collaborative research field of academic investigators - for subsequent endeavors. As more and more collaborative and interdisciplinary creations are made, an intention on the results regarding authorship tendencies and editorial group considerations for appropriate dissertation and article development will be of great use without extended lengths of abstracts. Furthermore, a more cohesive effort in communication will bolster worldwide recognition of articles by Bangladeshi scholars.

Conclusion

This research confirms that research team size is a significant predictor of Bangladeshi authors' language features of abstracts. With the presumption of larger collaborative teams being the most common collaborative authorship approach to what's currently out there, it seems that excessively sized contributions lend to longer - and thus less readable - abstracts. Abstracts with more like 5–10, however, have the potential for greater conceptual cohesion although it's determined that abstract language is more similarly aligned. Readability, however, is even more nuanced - not only is it not linear with the size of the team, but a substantial readable average can be attributed to a specific size (6–10). Average sized contributions with these authors produced better figures in legibility and syllable count than single written and those coalesced through excessive collaboration (11+).

At the same time, a relative reduction in readability across the board suggests an international phenomenon as specialized subfield researchers grow more specialized, developing more intricate methodologies and discipline-specific lexicon; this subfield position makes sense within a relatively new national research context like Bangladesh as it suggests needing help to create clear scientific writing practices and understandable practices through workshops, editorial support, and institutional development.

Thus, these findings would be generalizable through scientific writing workshops and optimal collaborative authorship conclusions that would be supported by national policies surrounding research. Early-career researchers would benefit from scientific writing workshops that promote appropriate collaborative styles, giving all researchers from Bangladesh awareness to the pros and cons of authorship on collaborative papers/literature. Nationally, best practice promotion of publication through clear communication would make scientifically minded pieces from Bangladesh come across internationally under a better (or at least preferred, clearer) light.

Future studies would seek to determine intentions from collaborative writers as to this phenomenon, longitudinal cross-sections of years with full manuscript assessments for within-team comparisons over the same years accessed per team size, interdisciplinary differences for national comparisons or international variances on abstract findings. All would explore collaborative connections between scientific communication in broader contexts.

Declarations

Funding:
The authors received no financial support for the research, authorship, and/or publication of this article.

Conflicts of Interest / Competing Interests:
The authors declare that there are no conflicts of interest.

Ethics Approval:
Not applicable. No human participants, animals, or sensitive data were involved.

Consent to Participate:
Not applicable.

Consent for Publication:
Not applicable.

Availability of Data and Materials:
The dataset derived from Web of Science is available from the corresponding author upon reasonable request.

Code Availability:
R scripts used for text mining, readability analysis, and visualization are available from the authors upon request.

Authors' Contributions (CRediT):

- **S. Ravikumar:** Conceptualization; Methodology; Data Curation; Formal Analysis; Visualization; Writing – Original Draft.
- **M. Mezbah-ul-Islam:** Supervision; Validation; Writing – Review & Editing; Interpretation.
- **Chandrappa:** Writing – review & editing; **Resources;** Validation**,** Investigation; Formal analysis; Writing – Original Draft.

References

1. Abt, H. A. (2007). The frequencies of multinational papers in various sciences. *Scientometrics*, *72*(1), 105–115. https://doi.org/10.1007/s11192-007-1710-5
2. Adams, J. D., Black, G. C., Clemmons, J. R., & Stephan, P. E. (2005). Scientific teams and institutional collaborations: Evidence from U.S. universities, 1981–1999. *Research Policy*, *34*(3), 259–285. https://doi.org/10.1016/j.respol.2005.01.014
3. Angulo-Tuesta, A., & Santos, L. M. P. (2015). Evaluation of the impact of maternal and neonatal morbidity and mortality research funded by the Ministry of Health in Brazil. *Research Evaluation*,

24(1), 1–11. https://doi.org/10.1093/reseval/rvv022

4. Ante, L. (2022). The relationship between readability and scientific impact: Evidence from emerging technology discourses. *Journal of Informetrics*, *16*, 101252. https://doi.org/10.1016/j.joi.2022.101252
5. Bao, T., Zhao, Y., Mao, J., & Zhang, C. (2025). Examining linguistic shifts in academic writing before and after the launch of ChatGPT: A study on preprint papers. *Scientometrics*. https://doi.org/10.1007/s11192-025-05341-y
6. Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, *3*(30), 774. https://doi.org/10.21105/joss.00774
7. Boyack, K. W., van Eck, N. J., Colavizza, G., & Waltman, L. (2018). Characterizing in-text citations in scientific articles: A large-scale analysis. *Journal of Informetrics*, *12*(1), 59–73. https://doi.org/10.1016/j.joi.2017.11.005
8. Bu, Y., Ding, Y., Liang, X., & Murray, D. (2018). Understanding persistent scientific collaboration. *Journal of the Association for Information Science and Technology*, *69*(4), 480–493. https://doi.org/10.1002/asi.23966
9. Chen, C. (2012). Predictive effects of structural variation on citation counts. *Journal of the Association for Information Science and Technology*, *63*(3), 431–449. https://doi.org/10.1002/asi.21694
10. Deng, H., Wu, C., Ge, B., & Wu, H. (2025). Quantification and identification of authorial writing style through higher-order text network modeling. *Journal of Informetrics*. https://doi.org/10.1016/j.joi.2024.101603
11. Dong, S., Mao, J., Ke, Q., & Pei, L. (2024). Decoding the writing styles of disciplines: A large-scale quantitative analysis. *Information Processing & Management*, *61*, 103718. https://doi.org/10.1016/j.ipm.2024.103718
12. Dressler, W., et al. (2025). Textual complexity and readability in scientific communication: A cognitive-linguistic perspective. *Journal of Technical Writing and Communication*, *55*(3), 413–439.
13. Franceschet, M., & Costantini, A. (2010). The effect of scholar collaboration on impact and quality of academic papers. *Scientometrics*, *87*(2), 295–306. https://doi.org/10.1007/s11192-010-0310-4
14. Gazni, A., Sugimoto, C. R., & Didegah, F. (2011). Mapping world scientific collaboration: Authors, institutions, and countries. *Journal of the Association for Information Science and Technology*, *63*(2), 323–335. https://doi.org/10.1002/asi.21688
15. Gazni, A., & Didegah, F. (2011). Investigating different types of research collaboration and citation impact: A case study of Harvard University publications. *Scientometrics*, *87*(2), 251–265. https://doi.org/10.1007/s11192-010-0318-9
16. Gui, W., Wang, H., Sun, X., & Wang, N. (2022). Linguistic complexity in scientific writing: A large-scale diachronic study from 1821 to 1920. *Scientometrics*, *127*(8), 4327–4348. https://doi.org/10.1007/s11192-022-04550-z

17. Hasegawa, Y., et al. (2024). A cell competition system with one gene expression from a single-copy gene in one cell. *PLOS ONE*, *19*, e302451. https://doi.org/10.1371/journal.pone.0302451
18. Hayati, Z., & Didegah, F. (2010). International scientific collaboration of Iran with other countries: A scientometric study. *Journal of Information Science*, *36*(4), 482–491. https://doi.org/10.1177/0165551509351430
19. Hirsch, J. E. (2010). An index to quantify an individual's scientific research output that takes into account the effect of multiple coauthorship. *Scientometrics*, *85*(3), 741–754. https://doi.org/10.1007/s11192-010-0193-9
20. Hoffman, E., Li, Y., Coutts, R. T., & Maris, S. (2024). Leveraging AI for educational content consistency: Metadata generation for MOOC development. *In IEEE DEMOcon 2024*. https://doi.org/10.1109/DEMOCON63027.2024.10747877
21. Lang, S., Walter, T., & Rodrigues, R. (2025). Readability and jargon in scientific plain-language summaries. *PLOS ONE*, *20*, e0302451.
22. Lancho-Barrantes, B. S., Guerrero-Bote, V. P., Chinchilla-Rodríguez, Z., & de Moya-Anegón, F. (2012). Citation flows in the zones of influence of scientific collaborations. *Journal of the Association for Information Science and Technology*, *63*(3), 481–489. https://doi.org/10.1002/asi.21682
23. Larivière, V., Sugimoto, C. R., & Gingras, Y. (2014). Team size matters: Collaboration and scientific impact since 1900. *Journal of the Association for Information Science and Technology*, *65*(4), 1323–1332. https://doi.org/10.1002/asi.23266
24. Lu, C., Bu, Y., Wang, J., Ding, Y., Torvik, V., Schnaars, M., & Zhang, C. (2019). Examining scientific writing styles from the perspective of linguistic complexity. *Journal of the Association for Information Science and Technology*, *70*(5), 462–475. https://doi.org/10.1002/asi.24126
25. Lu, C., Bu, Y., Dong, X., Wang, J., Ding, Y., Larivière, V., Sugimoto, C. R., Paul, L., & Zhang, C. (2019). Analyzing linguistic complexity and scientific impact. *Journal of Informetrics*, *13*, 100961. https://doi.org/10.1016/j.joi.2019.07.004
26. Lu, C., Zhang, C., Xiao, C., & Ding, Y. (2022). Contributorship in scientific collaborations: Contribution-based byline orders. *Information Processing & Management*, *59*, 102944. https://doi.org/10.1016/j.ipm.2022.102944
27. Lu, C., Zhang, Y., Ahn, Y.-Y., Ding, Y., Zhang, C., & Ma, D. (2020). Co-contributorship network and division of labor in individual scientific collaborations. *Journal of the Association for Information Science and Technology*, *71*(1), 78–92. https://doi.org/10.1002/asi.24321
28. Mayukh Bagchi. (2025). Toward generative AI–driven metadata modeling: A human–LLM collaborative approach. *Library Trends*. https://doi.org/10.1353/lib.2025.a961196
29. Mongeon, P., & Larivière, V. (2015). The career impact of research fraud: A scientometric analysis. *Journal of the Association for Information Science and Technology*, *66*(2), 205–217. https://doi.org/10.1002/asi.23202
30. Pan, X., Yan, E., Wang, Q., & Hua, W. (2015). Assessing the impact of software on science: A bootstrapped learning of software entities in full-text papers. *Journal of Informetrics*, *9*(2), 247–257.

https://doi.org/10.1016/j.joi.2015.07.012

31. Pinfield, S., Salter, J., Bath, P. A., Hubbard, B., Millington, P., Anders, J. H. S., & Hussain, A. (2014). Open-access repositories worldwide, 2005–2012. *Journal of the Association for Information Science and Technology*, *65*(*2*), 240–256. https://doi.org/10.1002/asi.23131
32. 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
33. Porwal, P., & Devare, M. (2024). Scientific impact analysis: Unraveling the link between linguistic properties and citations. *Journal of Informetrics*, *18*, 101526. https://doi.org/10.1016/j.joi.2024.101526
34. Romanchuk, R., Vysotska, V., Andrunyk, V., Chyrun, L., Chyrun, S., & Brodyak, O. (2023). Intellectual analysis system for Ukrainian-language artistic works. *In Proceedings of CSIT 2023*. https://doi.org/10.1109/CSIT61576.2023.10324012
35. Song, N., Chen, K., & Zhao, Y. (2023). Understanding writing styles of scientific papers in IS–LS: Evidence from abstracts. *Journal of Informetrics*, *17*, 101377. https://doi.org/10.1016/j.joi.2023.101377
36. Stephen, D. (2025). Distinguishing articles in questionable and non-questionable psychology journals. *Journal of Informetrics*, *19*, 101640. https://doi.org/10.1016/j.joi.2025.101640
37. Sun, Z., He, D., & Li, Y. (2024). How the readability of manuscripts influences peer review: Biomedical evidence. *Journal of Informetrics*, *18*, 101547. https://doi.org/10.1016/j.joi.2024.101547
38. Winnink, J., Tijssen, R., & Van Raan, A. F. J. (2015). Small team breakthroughs: Collaboration constellations in high-impact science. *Research Evaluation*, *24*(3), 352–366. https://doi.org/10.1093/reseval/rvv026
39. Wu, L., Wang, D., & Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. *Nature*, *566*(7744), 378–382. https://doi.org/10.1038/s41586-019-0941-9
40. Xu, S., Li, L., An, X., Hao, L., & Yang, G. (2021). Detecting commonality and specialty between scientific publications and patents. *Scientometrics*, *126*(3), 1789–1806. https://doi.org/10.1007/s11192-021-04085-9

Figures

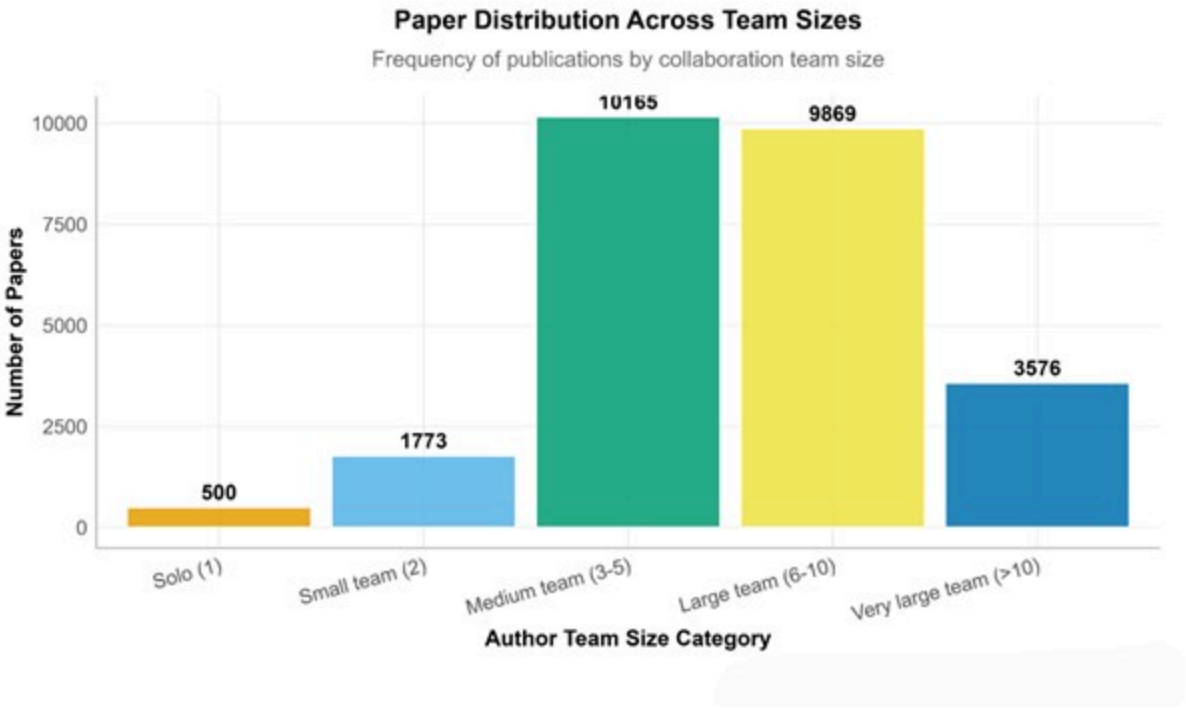


Figure 1

Frequency of publications by author team size category.

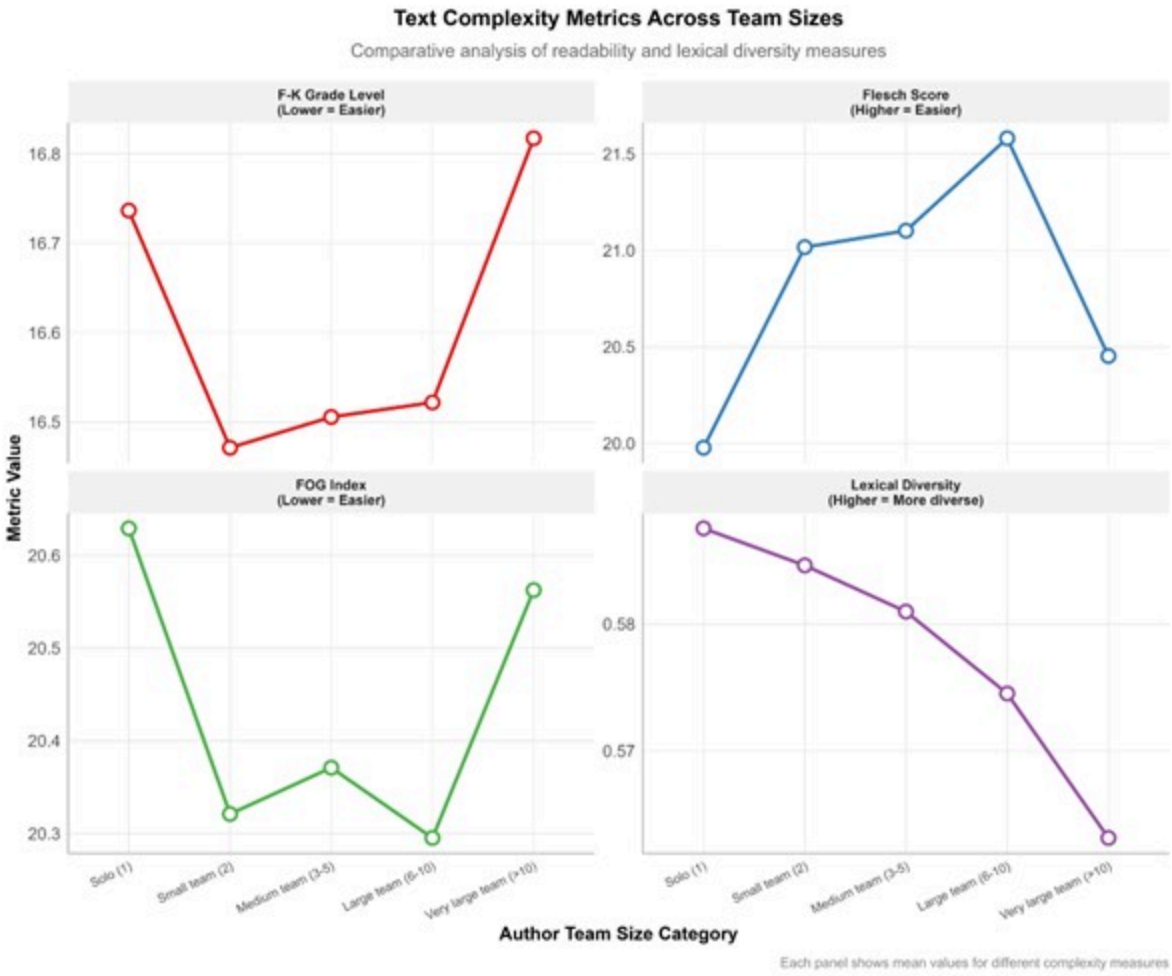


Figure 2

Mean values for readability and lexical diversity metrics by team size.

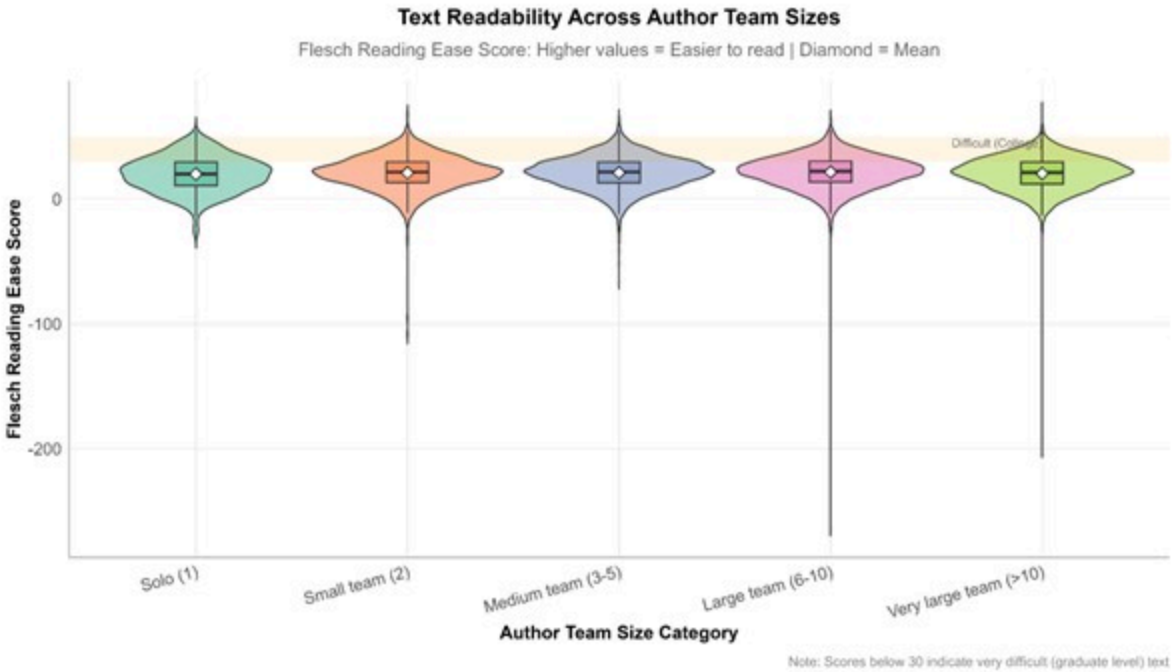


Figure 3

Violin plots of Flesch Reading Ease across team-size categories (diamond = mean).

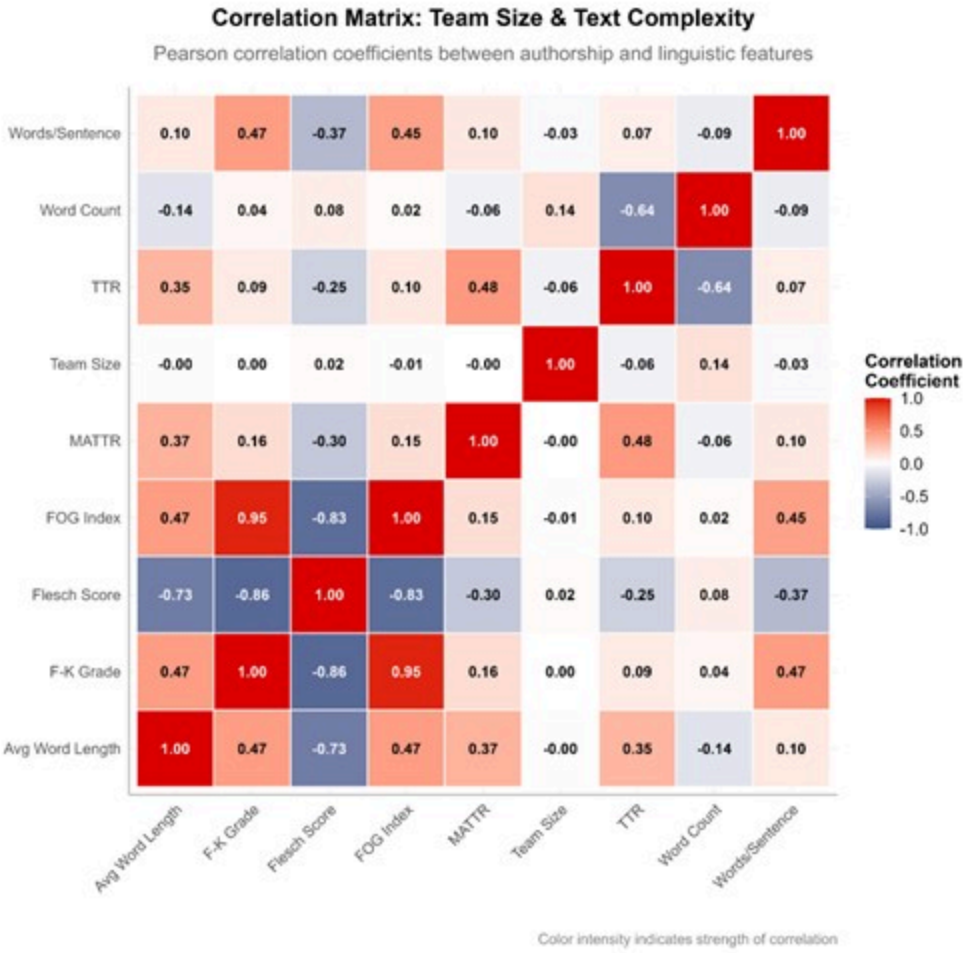


Figure 4

Correlation matrix of authorship and linguistic features.

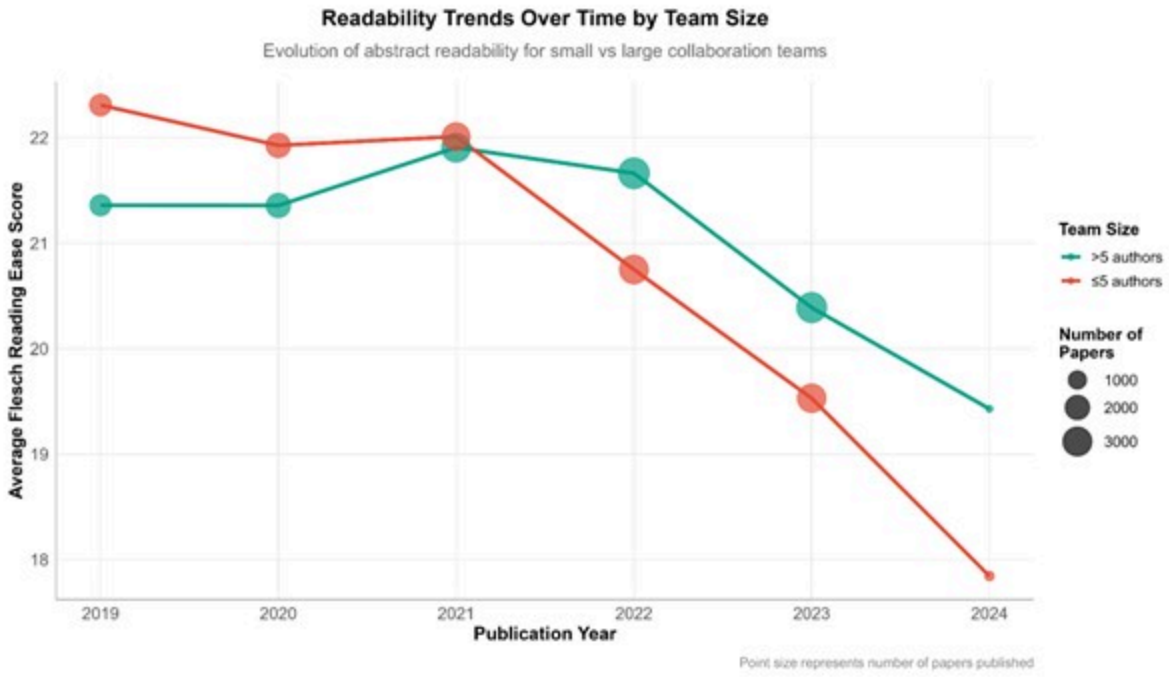


Figure 5

Evolution of abstract readability for small vs. large collaboration teams from 2019 to 2024.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Appendix.png