Peer Review

Ranjan, Ritvik, and Vidya Sinha. 2025. “A Multi-Dimensional Analysis of Linguistic Differences in Scientific Communication across Audiences and Disciplines.” *Journal of High School Science* 9 (4): 1–29. https://doi.org/10.64336/001c.145850

I enjoyed reading this manuscript. It is well presented and researched and does contribute to the corpus of existing knowledge in the field.
I do have some comments that need to be addressed. In addition, I am intrigued by the possibility of you introducing or suggesting an 'Exaggeration Index (EI) or an “Out Of Context” Index (OOSI) or a “Content dissimilarity index” (CDI) based on your lay versus technical compatison of WAGs (see point 4 below as well as the attached file).
1.You mention that you examined for normality. Did you perform a Levines test for homoskedasticity or equivalent? Please present in the manuscript.

2.Would it be possible to perform an ANOVA on the four disciplines (biological, physical, social) then followed by a Tuckey HSD post-hoc p-value analysis? Why was this statistical route not employed ? Please explain and describe in the manuscript.

3.I was hoping to see some word-adjacency graphs (WAG) (even though they may be presented in GitHub). Could you include some representative technical versus lay WAG in your manuscript with adequate descriptors of the various factors that you analyzed? I think those will be helpful to the reader, without having to navigate outside the manuscript.

4.I have frequently found that Lay descriptors exaggerate findings of the study, partially be taking them out-of-context. This may align with your larger lay ‘local-clusters’ as opposed to technical ‘universal-connections’ motifs. Would it be possible to construct an ‘exaggeration’ or ‘out-of-context’ (OOC) index using these differences ? You would need to correlate the ratio of (lay local clusters) to the (technical universal clusters) to the actual number or exaggerated words or simplistic descriptions encountered in the lay descriptions. From my experience as a bio-tech investor, I can personally tell you that this knowledge will be extremely helpful to investors; who can examine this ratio and decide how much of the “management summary” or the “10-K” reports represents an exaggeration (or is dissimilar to) of the actual findings of any particular scientific platform/publications. I will provide you with a simple example: The techincal summary may state " These findings are suggestive of an increased expression of the Alzheimer Neuro-fibriallary tangle reducing protein X upon the administration of drug Y, in-part due to the release of transcription factor Z from mutated complexasome A. This release of transcription factor Z may be triggered - in part - due to competion of binding sites on the mutated complexasome by drug Y, thereby blocking binding of transcription factor Z to the mutated inhibitory compexasome A. Therefore, the effect of the drug Y may be conditional upon a mutation in the binding pocket of complexasome A, limiting its potential to patients who exhibit this mutation. Because the majority of patients in this clinical trial exhibited this mutation in complexasome A, the phase III clinical trial was successful" Whereas, the lay summary may state: " In a major achievement, drug Y was found to be effective against Alzheimer’s disease in a phase III clinical trial".
I have also analyzed one of your abstracts from the GitHub site (see attached doc). The OOCI or the EI may indicate to what extent, this oversimplification by the lay-language distorts/exaggertes or takes content out of context (whether intentionally or intentionally) from the technical-language. You will need to manually curate all the 300 manuscript abstracts for this, but I think this is a worthwhile effort - considering that nothing like this yet exists in the literature.

5.How were those 300 articles chosen? Please describe randomization procedure. Were an equal number of articles chosen from each of the fields (biological, social and physical)? If not, why not? Implications?

6.I am assuming you are using IQR either because the data is skewed or you have outliers. In order to justify IQR, the # of 'outlier' data points should be below a certain percentage of the # of all the data points. Please discuss in the manuscript. A representative box-whisker plot with your worst ratio may be helpful in this regard.

7.References should be numbered sequentially in the text. The reference section should contain numbered references corresponding to those in the text. Please use APA format for reference formatting.

---

1 & 2. You mention that you examined for normality. Did you perform a Levines test for homoskedasticity or equivalent? Please present in the manuscript.
Would it be possible to perform an ANOVA on the four disciplines (biological, physical, social) then followed by a Tuckey HSD post-hoc p-value analysis? Why was this statistical route not employed ? Please explain and describe in the manuscript.

In addition to examining normality, we conducted Levene's tests across disciplines for each of the measures of complexity. These revealed significant violations for several key metrics (all $p < 0.05$). Because ANOVA assumes homoskedasticity, such violations would undermine its validity in our case.

By contrast, paired t-tests operate on within-item difference scores rather than raw group variances, making them robust to heteroskedasticity. Moreover, our research questions were focused on specific pairwise contrasts rather than an overall omnibus effect. For these reasons, we proceeded with paired t-tests rather than repeated-measures ANOVA with Tukey HSD post-hoc comparisons

3. I was hoping to see some word-adjacency graphs (WAG) (even though they may be presented in GitHub). Could you include some representative technical versus lay WAG in your manuscript with adequate descriptors of the various factors that you analyzed? I think those will be helpful to the reader, without having to navigate outside the manuscript.

Word Adjacency Graphs are included in the results section of the manuscript within the section on graph theoretic metrics. The sample graphs of an article's lay and technical summaries, created using MatPlotLib, are shown above a table that compares their graph theoretic metrics.

4. I have frequently found that Lay descriptors exaggerate findings of the study, partially be taking them out-of-context. This may align with your larger lay 'local-clusters' as opposed to technical 'universal-connections' motifs. Would it be possible to construct an 'exaggeration' or 'out-of-context' (OOC) index using these differences ? You would need to correlate the ratio of (lay local clusters) to the (technical universal clusters) to the actual number or exaggerated words or simplistic descriptions encountered in the lay descriptions. From my experience as a bio-tech investor, I can personally tell you that this knowledge will be extremely helpful to investors; who can examine this ratio and decide how much of the "management summary" or the "10-K" reports represents an exaggeration (or is dissimilar to) of the actual findings of any particular scientific platform/publications. I will provide you with a simple example: The techincal summary may state " These findings are suggestive of an increased expression of the Alzheimer Neuro-fibriallary tangle reducing protein X upon the administration of drug Y, in-part due to the release of transcription factor Z from mutated complexasome A. This release of transcription factor Z may be triggered - in

part - due to competion of binding sites on the mutated complexasome by drug Y, thereby blocking binding of transcription factor Z to the mutated inhibitory compexasome A. Therefore, the effect of the drug Y may be conditional upon a mutation in the binding pocket of complexasome A, limiting its potential to patients who exhibit this mutation. Because the majority of patients in this clinical trial exhibited this mutation in complexasome A, the phase III clinical trial was successful" Whereas, the lay summary may state: " In a major achievement, drug Y was found to be effective against Alzheimer's disease in a phase III clinical trial".
I have also analyzed one of your abstracts from the GitHub site (see attached doc). The OOCI or the EI may indicate to what extent, this oversimplification by the lay-language distorts/exaggertes or takes content out of context (whether intentionally or intentionally) from the technical-language. You will need to manually curate all the 300 manuscript abstracts for this, but I think this is a worthwhile effort - considering that nothing like this yet exists in the literature.

We appreciate the reviewer's insightful and thought-provoking suggestion to create an "Exaggeration" or "Out-of-Context" Index (OOCI). We agree that such a tool would be a valuable contribution to a variety of audiences, such as investors.

While we find this idea compelling and a potential avenue for future research, we have decided not to pursue it in this paper. Our study is focused on a multi-dimensional linguistic analysis using quantifiable, automated metrics to compare lay and technical abstracts. Our methodology is built on:

- Syntactic complexity measures (e.g., MLS, DC/T) from Lu's (2010) L2SCA.
- Graph-theoretic metrics (e.g., Average Nodal Degree, Graph Density) to analyze word co-occurrence.

Developing an OOCI would require a qualitative, content-based assessment of each paper's findings, moving beyond our current computational framework. This would necessitate a fundamentally different research design and would be more suitable for a dedicated, separate study.

We believe that incorporating this qualitative analysis would move the paper away from its central research questions about linguistic and structural complexities and into a distinct area of content analysis. We feel that this would detract from the cohesive focus of our current manuscript.

Thank you again for this excellent suggestion, which has provided us with a new direction to consider for future research.

5. How were those 300 articles chosen? Please describe randomization procedure. Were an equal number of articles chosen from each of the fields (biological, social and physical)? If not, why not? Implications?
The 300 articles were chosen for their viewership in order to ensure the works we were analyzing were public-facing, ensuring the corpus reflected research with the broadest public interest and relevance. Thus we had representations of each field that, while slightly unequal, represented broader societal trends of paper publishing volume, such as the prevalence of publicly relevant biology articles. Although the representation across fields was uneven, this did not influence our

findings because our comparisons were not dependent on raw article counts. This explanation can be found in our methodology section, under “Corpus and Data collection.”

6. I am assuming you are using IQR either because the data is skewed or you have outliers. In order to justify IQR, the # of ‘outlier’ data points should be below a certain percentage of the # of all the data points. Please discuss in the manuscript. A representative box-whisker plot with your worst ratio may be helpful in this regard.

For each complexity measure and subject, we verified that the proportion of outliers did not exceed 25% of the total data points, ensuring that the interquartile range (IQR) provided a stable and representative measure of variability. This table is featured in the methods section of the manuscript.

	Worst Ratio (outliers / total data points)
Syntactic complexity measures for lay summaries	0.1
Syntactic complexity measures for technical summaries	0.03
Graphical complexity measures for lay summaries	0.04
Graphical complexity measures for technical summaries	0.06

7. References should be numbered sequentially in the text. The reference section should contain numbered references corresponding to those in the text. Please use APA format for reference formatting.
We adjusted the citations throughout the paper as instructed.

Thank you for addressing my comments.

1.I am happy with most of them except my comment about the out-of-context index. I will buy your explanation that this would require much more work. However, that should not prevent you from describing this in a ‘Perspectives’ section of the manuscript. Please do so. You can even include my biotech example or include your own.

2.I recommend that you remove all the rows in all the tables for which the p-values are > 0.05. They serve no purpose and only distract from the salient features.

3.For those rows for which you have significant differences, I would like to see more explanation on why those differences occur in some comparative areas but not in others. For example, for Table 1, why are only MLS, MLT, MLC…… different between lay and technical summaries AND NOT THE OTHER (attributes in the first column). Similarly, for Table 3, why does MLS only show a statistical difference between physical versus biological but NOT between physical versus social or biological versus social…… present this analysis for all tables and signficant rows. This will add

more depth to your manuscript. Put this down in the discussion section. You can even add another Table to the manuscript with descriptors such as "Reason(s) why significant column if you so choose.

4.Present the Summary as two paragraphs. The Table format is unweildy for this.

---

1. I am happy with most of them except my comment about the out-of-context index. I will buy your explanation that this would require much more work. However, that should not prevent you from describing this in a 'Perspectives' section of the manuscript. Please do so. You can even include my biotech example or include your own.

We added a "Perspectives" section below our Conclusion to address the idea. We included the biotech example and described the framework for such a project.

2. I recommend that you remove all the rows in all the tables for which the p-values are > 0.05. They serve no purpose and only distract from the salient features.

For the tables that did not depict cross-disciplinary differences, we removed the rows with p-values > 0.05.
For tables that exhibited multiple p-values associated with the three disciplinary umbrellas, we removed the rows that had no p-values < 0.05 and preserved the rows that had at least one p-value < 0.05, because we did not want to omit any significant values. We expressed this omission of non-significant metrics in the "Data Reporting" sub-section within the Discussion section.

3. For those rows for which you have significant differences, I would like to see more explanation on why those differences occur in some comparative areas but not in others. For example, for Table 1, why are only MLS, MLT, MLC…… different between lay and technical summaries AND NOT THE OTHER (attributes in the first column).
Similarly, for Table 3, why does MLS only show a statistical difference between physical versus biological but NOTbetween physical versus social or biological versus social…… present this analysis for all tables and signficantrows. This will add more depth to your manuscript. Put this down in the discussion section. You can even addanother Table to the manuscript with descriptors such as "Reason(s) why significant column if you so choose.

We added an opening section to the discussion highlighting the similarities between lay and technical summaries, focusing on metrics that showed no significant differences and the implications of these shared linguistic foundations, and incorporated two tables that explain why specific metrics do not display significant differences between disciplines in both lay and technical summaries.

4. Present the Summary as two paragraphs. The Table format is unweildy for this.
We implemented this change, moving the summaries above images 1 and 2 within the results section.

Thank you for addressing my comments. Accepted.