

Revealing the Hidden Success Factors of Scientific Papers: The Case of Table and Figure References.

Fabian Haak^a, Philipp Schaer^a

^a TH Köln University of Applied Sciences, 54 Gustav-Heinemann-Ufer, Cologne, 50968, Germany

Authors are provided with guidelines and style manuals designed to make their scientific manuscripts “successful.” These guides often recommend specific linguistic tones, strategic placement of illustrative figures, or avoiding repetitive information between text and tables. But do these structural and stylistic choices actually influence a paper’s impact? Utilizing a large dataset of parsed life science preprints, we investigate the usage of tables, figures, their captions, and in-text references. By analyzing linguistic patterns, structural complexity, and version histories, we empirically assess whether adherence to established author guidelines correlates with actual scientific success.

Authors of scientific papers have many guidelines and style guides at hand to maximize the impact and “success” of their work (Budgell, 2008). These manuals frequently suggest practices such as using a specific type of language, placing figures at the very beginning to grab attention, or ensuring that the main text does not simply repeat data already presented in tables (Zhang, 2014).

However, a critical question remains largely unanswered: Do these guidelines really help and make a paper successful? Are these structural and stylistic choices genuine drivers of scientific impact? To answer this, we move beyond intuition. Multimodal information, specifically tables, figures, and their associated textual anchors (captions and in-text references), shapes how a paper is comprehended. This study aims to empirically validate these assumed best practices by analyzing how authors integrate non-textual elements and whether these structural choices serve as hidden indicators of scientific success. We analyze a large corpus of parsed life science preprints from the bioRxiv platform². This dataset allows us to examine not only the static structure of a paper but also the evolutionary changes made across multiple manuscript versions prior to and post formal publication. Using the Reanimator framework (Engelmann et al., 2025), we parse full texts, tables, and figures, as well as captions and corresponding in-text references. We then map these elements and their structural traits against established metrics of scientific success, such as the eventual peer-reviewed publication venue and community impact (e.g., citations, impact factor). We investigate how structural choices impact both human comprehension and machine discoverability, e.g., for search engines and RAG. Specifically, we investigate:

- What specific information is conveyed in tables but omitted from the main text, alongside the linguistic heuristics used to condense tabular data into continuous text?
- We analyze the specific linguistic patterns authors use to compress deep, multi-dimensional visual data into localized text blocks.
- By comparing distinct preprint versions, we identify the typical structural and textual edits made to tables and figures during the revision lifecycle.
- How closely do manuscripts align with common author guidelines, and how does this adherence drive greater impact?

This poster presentation will provide preliminary results from a large-scale empirical study regarding the structural mechanics of successful scientific communication. By revealing which multimodal guidelines genuinely matter, our findings will offer actionable insights for researchers

SEASON 2026, September 15–17, 2026, Hamburg, Germany
EMAIL: fabian.haak@th-koeln.de (F. Haak); philipp.schaer@th-koeln.de (P. Schaer)
ORCID: [0000-0002-0877-7063](https://orcid.org/0000-0002-0877-7063) (F. Haak); [0000-0002-9421-8566](https://orcid.org/0000-0002-9421-8566) (P. Schaer)
DOI: <https://doi.org/10.48441/4427.3736>



© 2026 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

² <https://biorxiv.org/>



aiming to optimize their manuscripts and for publishers seeking to refine their editorial standards based on data rather than tradition.

1. References

B. Budgell, Writing a Biomedical Research Paper: A Guide to Structure and Style, Springer Science & Business Media, 2008.

W. Zhang, Ten Simple Rules for Writing Research Papers, PLoS Computational Biology 10 (2014) e1003453. doi:10.1371/journal.pcbi.1003453.

B. Engelmann, F. Haak, P. Schaer, M. Erfanian Abdoust, L. Netze, M. Bittkowski, Reanimator: Reanimate retrieval test collections with extracted and synthetic resources, in: Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '25, Association for Computing Machinery, New York, NY, USA, 2025, p. 3691–3701. URL: <https://doi.org/10.1145/3726302.3730342>. doi: 10.1145/3726302.3730342.