Principles of Scientific Writing

Douglas Ezra Morrison

2025-09-19

Table of contents

Preface

This book will present my perspective on scientific writing, in particular focusing on writing about statistical data analysis, but also scientific writing generally (and informative prose generally). I intend to give it to academic mentees as a style guide reference, and also use it as a notebook for myself.

1 Introduction

2 Defining terms clearly

Clear definitions are essential to effective scientific writing. Every specialized term should have an explicit, concise definition immediately before or after its first use. Readers should not need to search for the meaning of a term or infer it from context alone. For example:

Definition 2.1 (Term)**.** A *term* is a word or phrase with a specific, technical meaning within a particular field or context.

Example 2.1 (Population)**.**

Definition 2.2 (Population)**.** In statistics, the term “population” refers to the complete set of all items or individuals of interest, not just a geographic population of people.

2.1 Guidelines for defining terms

Follow these principles when introducing new terms:

- **Provide explicit definitions:** State clearly what a term means, rather than assuming readers will understand from context.
- **Be concise:** Definitions should be brief and focused, using only the words necessary to convey the meaning.
- **Define terms at first use:** Place the definition immediately before or after the term’s first appearance.
- **Provide examples:** Every definition should include at least one concrete example that illustrates how the term is used.

2.2 Examples of term definitions

Here are examples of well-defined terms:

Example 2.2 (Statistical term: Confidence interval)**.**

Definition 2.3 (Confidence interval)**.** A *confidence interval* is a range of values, derived from sample statistics, that is likely to contain the true population parameter.

If we calculate a 95% confidence interval for mean height as [165 cm, 175 cm], we are 95% confident that the true population mean height falls within this range.

Example 2.3 (Writing term: Active voice)**.**

Definition 2.4 (Active voice)**.** *Active voice* is a sentence structure where the subject performs the action expressed by the verb.

“The researcher conducted the experiment” is in active voice, while “The experiment was conducted by the researcher” is in passive voice.

2.3 Why examples matter

Examples serve several crucial purposes:

- **Concrete understanding:** Examples transform abstract concepts into tangible instances that readers can visualize and understand.
- **Verification:** Readers can work through examples themselves to verify they understand the definition or theorem.
- **Application:** Examples show how to apply definitions and theorems to solve real problems.
- **Memory:** Concrete examples are easier to remember than abstract definitions alone.

2.4 Common pitfalls to avoid

- **Circular definitions:** Do not define a term using the term itself.

  Incorrect: "Conciseness is the quality of being concise."

  Correct: "Conciseness is the quality of expressing ideas using only the words necessary to communicate meaning clearly."

- **Vague definitions:** Avoid definitions that rely on imprecise language.

  Incorrect: "A variable is something that can change."

  Correct: "A variable is a symbol representing a value that can take on different values."

- **Missing examples:** Never leave a definition without at least one example.

- **Examples without context:** Ensure examples clearly illustrate the specific aspect of the definition they are meant to demonstrate.

2.5 Mathematical statements require examples

Mathematical theorems, lemmas, corollaries, and postulates should always include examples that demonstrate their application. Abstract mathematical statements become more accessible when readers can see concrete instances.

2.5.1 Using theorem environments in Quarto

When writing mathematical or technical content in Quarto, use theorem environments to demarcate and highlight the structure of your content. These environments provide consistent formatting, automatic numbering, and cross-reference capabilities.

Quarto provides built-in theorem environments including:

- `#thm-` for theorems
- `#lem-` for lemmas
- `#cor-` for corollaries
- `#prp-` for propositions
- `#cnj-` for conjectures
- `#def-` for definitions
- `#exm-` for examples
- `#exr-` for exercises

See the Quarto documentation on theorems and proofs for complete details.

Example syntax:

```
::: {#thm-pythagorean}

## Pythagorean theorem

In a right triangle,
the square of the hypotenuse equals the sum of squares of the other two sides:
$a^2 + b^2 = c^2$.

:::
```

This produces automatically numbered output like “Theorem 2.1 (Pythagorean theorem)” and can be referenced elsewhere in your document using `@thm-pythagorean`.

2.5.2 Example: Pythagorean theorem

Theorem 2.1 (Pythagorean theorem)**.** *In a right triangle, the square of the hypotenuse equals the sum of squares of the other two sides:* $a^2 + b^2 = c^2$*.*

Example 2.4 (Pythagorean theorem example)**.** For a triangle with sides 3, 4, and 5: $3^2 + 4^2 = 9 + 16 = 25 = 5^2$.

Definition 2.5 (Postulate (axiom))**.** A *postulate* (or axiom) is a statement accepted as true without proof, serving as a starting point for further reasoning.

Example 2.5 (Euclid’s parallel postulate)**.** Through a point not on a line, exactly one line can be drawn parallel to the given line.

If we have line L and point P not on L, only one line through P will never intersect L.

3 Citations and Evidence

Every claim in scientific writing should be supported by either citations to relevant sources or direct evidence from data or experiments. This principle is fundamental to maintaining credibility, enabling verification, and building on the accumulated knowledge of the scientific community.

3.1 Why claims need support

Scientific writing aims to convey truthful, verifiable information. Unsupported claims undermine this goal in several ways:

- **Credibility**: Readers cannot assess the reliability of assertions without knowing their basis
- **Verification**: Claims without support cannot be independently verified or challenged
- **Reproducibility**: Other researchers need sources to understand the foundation of your work
- **Intellectual honesty**: Proper attribution gives credit where it is due and distinguishes your original contributions from existing knowledge
- **Learning pathway**: Citations provide readers a roadmap to learn more about a topic

When you make a claim, you are asking readers to accept it as true. They deserve to know why they should trust that claim.

3.2 What constitutes adequate support

Different types of claims require different types of support:

3.2.1 Empirical claims

Claims about observable phenomena or data should be supported by:

- Direct presentation of the data (tables, figures, statistical analyses)
- Citations to published studies that collected the relevant data

- Description of experimental methods that generated the evidence
- Links to publicly accessible datasets

3.2.2 Theoretical claims

Claims about concepts, models, or interpretations should be supported by:

- Citations to papers that developed or validated the theory
- Logical arguments with clearly stated premises
- Mathematical derivations or proofs
- References to review articles that synthesize relevant theoretical work

3.2.3 Methodological claims

Claims about appropriate methods or best practices should be supported by:

- Citations to methodological papers or textbooks
- Empirical evidence of method performance
- Expert consensus statements
- Validation studies

3.2.4 Common knowledge

Not every statement requires citation. Well-established facts that are common knowledge within your field (e.g., "DNA is a double helix" in molecular biology) can be stated without citation. However, when in doubt, provide a citation—over-citing is preferable to under-citing.

3.3 What makes a citation relevant

A relevant citation is one that actually supports the specific claim you are making. Common problems with citation relevance include:

- **Citing review papers indiscriminately**: While review papers are valuable for general background, cite the original research when making specific claims about particular findings
- **Citing papers that don't address the claim**: Ensure the cited work actually discusses the point you're making
- **Over-generalizing from narrow studies**: Don't cite a study with limited scope to support a broad general claim

- **Using outdated sources**: When current knowledge has superseded older findings, cite more recent work

To ensure relevance:

- Read the papers you cite (at least the relevant sections)
- Verify that the cited work actually supports your specific claim
- Cite the most direct source available
- When citing for general background versus specific claims, make the distinction clear

3.4 What makes a source trustworthy

Not all sources are equally reliable. Consider these factors when evaluating trustworthiness:

3.4.1 Peer review

Peer-reviewed publications in reputable journals have undergone expert scrutiny. This doesn't guarantee correctness, but it provides a baseline level of quality control.

3.4.2 Reputation of authors and institutions

Work from recognized experts and well-regarded institutions tends to be more reliable, though this should not be the sole criterion.

3.4.3 Replication and consensus

Findings that have been replicated by independent groups or that represent scientific consensus are more trustworthy than isolated claims.

3.4.4 Transparency and reproducibility

Studies that:

- Clearly describe their methods
- Share their data and code
- Disclose potential conflicts of interest
- Have been successfully reproduced by others

are more trustworthy than those lacking these features.

3.4.5 Preprints and non-peer-reviewed sources

Preprints can be valuable for accessing cutting-edge research, but they have not undergone peer review. When citing preprints:

- Note that they are preprints
- Check whether a peer-reviewed version has since been published
- Exercise extra scrutiny of the methods and conclusions

3.4.6 Sources to generally avoid

Some sources typically lack the rigor needed for scientific writing:

- Wikipedia and general encyclopedias (though they can be useful starting points for finding primary sources)
- Popular press articles about scientific findings (cite the original research instead)
- Blogs and social media posts (unless discussing public discourse or documenting specific claims made in those venues)
- Predatory or pay-to-publish journals without genuine peer review
- Retracted papers

3.5 Best practices

To effectively support your claims:

1. **Cite as you write**: Add citations immediately when making claims, rather than planning to "add references later"
2. **Use citation management tools**: Software like Zotero, Mendeley, or BibTeX helps organize and format references correctly
3. **Check your citations**: Before submitting, verify that every citation supports its associated claim
4. **Provide context**: Help readers understand why a source is relevant (e.g., "Smith et al. (2020) demonstrated that…" rather than just "(Smith et al., 2020)")
5. **Balance primary and review sources**: Use primary sources for specific findings, reviews for general background
6. **Stay current**: Supplement foundational older references with recent work showing the current state of knowledge
7. **Cite diverse sources**: When possible, include work from different research groups and perspectives

3.6 Common citation errors to avoid

- **Citation needed**: Making claims without any supporting citation or evidence
- **Vague attribution**: Using phrases like "studies have shown" without citing specific studies

- **Circular citation**: Citing a paper that doesn't contain the claimed information but cites another paper that does (cite the original source)
- **Citation padding**: Adding citations that don't actually support your claims just to appear well-referenced
- **Selective citation**: Only citing work that supports your position while ignoring contradictory evidence
- **Ghost authorship**: Failing to cite work that directly influenced your ideas

3.7 Examples

3.7.1 Poor (unsupported claim)

> Machine learning models often perform poorly on small datasets.

Problem: This claim is stated as fact without any support.

3.7.2 Better (citation provided)

> Machine learning models often perform poorly on small datasets (Vapnik 1998; Hawkins 2004).

Improvement: Citations provide evidence for the claim.

3.7.3 Best (citation with context)

> Machine learning models often perform poorly on small datasets. Vapnik (1998) showed that the generalization error of learning algorithms typically decreases as training set size increases, and Hawkins (2004) demonstrated that complex models are particularly prone to overfitting when trained on limited data (Vapnik 1998; Hawkins 2004).

Improvement: The specific support each citation provides is explained.

3.7.4 Poor (irrelevant citation)

> Python is the most popular programming language for data science (Knuth 1984).

Problem: Knuth's 1984 paper on literate programming doesn't address Python or data science.

3.7.5 Better (relevant citation)

> Python is the most popular programming language for data science (Stack Overflow 2024).

Improvement: The citation is to a current survey of programming language usage.

3.8 Conclusion

Supporting claims with appropriate citations and evidence is not optional—it is essential to scientific communication. It allows readers to verify your claims, understand the foundation of your arguments, and locate resources for further learning. Always ask yourself: "How does my reader know this is true?" If the answer isn't obvious from your text, add a citation or present direct evidence.

4 Word choice

I recommend trying to replace Latin-derived words and phrases with Old English-derived equivalents ("Anglish" words) where possible; it generally makes writing simpler and easier to read. Latin words create artificial barriers to understanding. Many Latin-derived words commonly used in scientific writing are composed from roots and affixes which are not commonly used in their basic forms; hence, readers cannot determine the meanings of these words by decomposing them. Instead, they need to memorize the meanings of these words directly. In contrast, the components of composite Anglish words are typically also used individually, so the meanings of the composites can be derived directly. Table 4.1 lists some common Latinate words and phrases and Anglish alternatives.

Table 4.1: Commonly used Latin words and phrases and Anglish alternatives

Latin	Anglish
prior to	before
necessary	needed

See also https://bark-fa.github.io/Anglish-Translator/

I am aware that this book, and even this chapter, contains many Latin word choices where there are Anglish alternatives. It is a work in progress, and also, I am not advocating 100% Anglish purity. Use whichever words and phrases you think your readers are most likely to understand easily. Preferring Anglish is merely a useful heuristic to help achieve our ultimate goal of producing clear, easy-to-read writing.

Just to be clear, although I prefer Anglish words, I have no particular preference for Anglish people or culture; it is only a practical consideration, based on the realities of English as the current default language of science and the relatively-recent hybridization of the English language.

5 Conciseness

Concise writing conveys ideas efficiently, using only the words necessary to communicate meaning clearly. Every word should serve a purpose. Eliminate redundancy and verbosity. When you can express an idea in fewer words without losing meaning, do so.

5.1 Common ways to improve conciseness

- Remove redundant phrases
  - “in order to” → “to”
  - “due to the fact that” → “because”
  - “at this point in time” → “now”
  - “a large number of” → “many”
- Use active voice instead of passive where appropriate
  - “The experiment was conducted by the researchers” → “The researchers conducted the experiment”
- Eliminate unnecessary qualifiers
  - “very”, “really”, “quite” often add little meaning
- Replace wordy phrases with single words
  - “make a decision” → “decide”
  - “give consideration to” → “consider”
  - “is able to” → “can”

Remember: concise writing is not about making every sentence as short as possible, but about removing words that do not contribute to meaning or clarity.

5.2 Examples of concise writing

Many effective writers throughout history have exemplified the principle of conciseness.

Julius Caesar's *Commentarii* and phrases like "Veni, vidi, vici" demonstrate the power of brevity (Wikipedia contributors 2026b).

Ernest Hemingway was known for his spare, direct prose style. His short sentences and simple words conveyed complex ideas and emotions without unnecessary embellishment (Wikipedia contributors 2026a).

6 Summary

In summary, this book has no content whatsoever.

References

Hawkins, Douglas M. 2004. "The Problem of Overfitting." *Journal of Chemical Information and Computer Sciences* 44 (1): 1–12. https://doi.org/10.1021/ci0342472.

Knuth, Donald E. 1984. "Literate Programming." *Comput. J.* 27 (2): 97–111. https://doi.org/10.1093/comjnl/27.2.97.

Stack Overflow. 2024. "Stack Overflow Developer Survey 2024." https://survey.stackoverflow.co/2024/.

Vapnik, Vladimir N. 1998. *Statistical Learning Theory*. New York: Wiley. https://www.wiley.com/en-us/Statistical+Learning+Theory-p-9780471030034.

Wikipedia contributors. 2026a. "Ernest Hemingway — Wikipedia, the Free Encyclopedia." https://en.wikipedia.org/wiki/Ernest_Hemingway#Writing_style.

———. 2026b. "Julius Caesar — Wikipedia, the Free Encyclopedia." https://en.wikipedia.org/wiki/Julius_Caesar#Literary_works.