Guide to Writing a Scientific Paper

Dean Wheeler
Brigham Young University
September 2020

This guide provides details and expectations for writing a scientific paper. The principles also apply to other writing such as a thesis or dissertation. At the end I give writing rules that I consider important.

1 Objectives and Audience

For any kind of writing you should understand the objectives of the document and tailor your material for the intended audience. The objectives for almost all technical writing are to

- **Sell your ideas**. This means you should convince the intended audience that your work and conclusions are valid, interesting, and useful.
- **Disseminate knowledge and teach**. This means you should accurately and concisely document your procedures, calculations, and results to allows others to verify and implement.

A scientific paper will describe

- An important problem that was solved or question that was answered or discovery that was made.
- The process you designed to solve that problem or answer that question. This could be through constructing a tool or apparatus, experimental or computational data collection, review of previously collected data, or combination of these.
- Details of the resulting data and analysis process, including appropriate use of statistics.
- Sound conclusions drawn from the data.

The **audience** is typically a subset of the engineering and scientific community specific to your discipline. Audience members will typically be familiar with common problems in the discipline and techniques to solve them. However, they have not typically been working on your specific problem or may be unfamiliar with your proposed approach and so a brief review of background material, terminology, and theory specific to the topic will be helpful. They need to be told what is new and innovative about your work.

2 Paper Content

A scientific manuscript is around 25 pages long (double space, including figures but not including bibliography and appendices). After formatting by the journal it will be about one-third this length. Below is the general order of sections, along with a suggested length for each. Some adaptation may be warranted depending on the nature of the project.

1. **Title and Authors**

   Choosing an appropriate title that is specific to your topic (but not too specific) will require some trial and error. Start with a few candidate titles and then narrow down to the one that works best.

   By convention, authors are listed in decreasing order of their overall contribution to the work and writing, with the exception of the senior author (i.e. main advisor or professor) who is generally listed at the end. The "corresponding author" is the one whose email address is listed and who should be contacted if readers have questions about the work.

2. **Abstract** (0.5 page maximum)

   The abstract contains an overview of the problem being solved, the tools or methods used to solve the problem, the principal results and conclusions, and the value or impact of the work. You are trying to sell the work: give sufficient information that readers can determine if they should spend time reading the main document. The abstract is independent and stands apart from the other sections, meaning everything in the abstract must be re-introduced and repeated in the main document. Generally citations are not given in the abstract unless the paper is a specific response to or continuation of a prior publication.

3. **Introduction** (2–3 pages)

   (a) Give a brief background of your field of work, identify a problem, and establish the problem's importance. This is your opportunity to generate interest in the audience about your project. Avoid using highly technical jargon at this point and don't make this part longer than 1.5 pages.

   (b) State the scope of your project, namely the problem you solved, question you answered, or hypothesis you tested. Summarize your approach to solve the problem: building an apparatus, conducting experiments, doing computations, analyzing previously collected data, or a combination of these.

   (c) Outline the topics that are covered in the remainder of the document, i.e. give a road map.

   The introduction has a parallel purpose to the abstract in securing reader interest and setting expectations for the remainder of the document. Therefore, make sure the introduction is particularly interesting, effective, and organized. Writing a good introduction is challenging and requires significant time and editing!

4. **Background** (3–5 pages, with a subsection for each topic)

   Do a more extensive background discussion on the problem, showing you have a good grasp of the field in which you are working. You are educating your audience to better understand the nature of the problem and appreciate the innovation of your solution.

   This section can incorporate your "literature review," though give it a more specific title than this. In particular, describe where others' prior work overlaps with or impacts your work, showing ideas you gained from them or knowledge gaps that existed prior to your work. You should reference at least 15 sources in the paper—some references are more important and will deserve individual and critical analysis, while other references can be lumped together as part of a discussion of related works that address a particular issue. Some references can be saved for, or revisited in, the Results section below in order to compare your outcomes with those of prior work.

5. **Methods** (around 3–9 pages, with a subsection for each topic)

   Give a detailed description of the experimental, mathematical, and computer methods and conditions used to generate results. Provide the logic guiding your design choices. Describe efforts to control uncertainty and validate that methods were working as intended. Provide figures and tables as needed. There should be sufficient detail that another scientist could duplicate your work. If you are using established methods described in prior publications, then use the smaller end of the page range; if the main point of the work is the development of a new apparatus or method, then use the longer end of the page range.

6. **Results and Discussion** (around 7–13 pages, with a subsection for each topic)

   This is the heart of the paper where you teach your audience what was learned and achieved from your project. Essentially you present a series of figures and tables, with each having one or more accompanying paragraphs describing the story told by the data. Figures can include images, conceptual diagrams, and plots of data. Whenever possible provide confidence intervals for each data point. Compare your results to those of prior work to explain differences and make connections.

7. **Conclusion** (1–2 pages)

   The conclusion should do more than just repeat or summarize the preceding material; it also should make clear the key insights and lessons from the work, draw attention to what was innovative, and express how the results might impact more broadly than the original problem. It could also suggest possible extensions to or planned continuation of the work.

8. **Bibliography**

   Give a list of at least 15 cited references, numbered in the order they first appeared in the main document. Use a consistent format taken from the journal to which you are submitting. It is common for bibliography entries to get misformatted or to be missing key information. Therefore, double check each reference.

9. **Appendix** or **Supplementary Material** (optional)

   An appendix or supplementary section is the best place to put extended data sets, lengthy mathematical derivations, or computer code that would otherwise interrupt or distract from the logical flow of the main part of the document or would be of less interest to the general audience. Some journals also allow you to upload video content.

3 Writing Rules

Below is my advice for improving your technical writing, in the form of 38 rules. They are based on my observations of trouble spots for beginning (and sometimes more-experienced) technical writers.

General Scientific Writing

1. Good writers habitually reread and rewrite what they have written to fix mistakes and make improvements. They take ownership of their work and pride in producing a high-quality product. The best writing requires many edits and drafts over the course of days and weeks. Printing out a draft of your document can be helpful in detecting mistakes, because the text and formatting will appear differently to your eye than on a computer screen: try it and see.

2. When someone else edits your document, learn from their corrections. Fix all instances of that mistake in the document and do not repeat the same mistake in the future. The editor may be your advisor or simply a "fresh pair of eyes," a person who doesn't necessarily need a technical background. There is a temptation to postpone fixing grammar and spelling mistakes. However, with a polished draft your advisor can spend more time addressing "big picture" things like the logical structure and flow of ideas, rather than getting distracted by the grammar and spelling mistakes.

3. Good writers are teachers. They tune their writing to their audience's general level of technical experience and prior knowledge and interest in the topic, which will be different from the writer's current knowledge and interest. Try to *anticipate* a typical reader's questions and objections, gaps in knowledge, and what they will find confusing or tedious. Then modify your writing to meet those needs. Similarly, follow the golden rule in writing: write the kind of paper that you like to read when you are new to a topic. Think about particular documents that have been especially helpful to you as a beginner. How did the authors teach you? Then write that kind of paper.

4. Scientific writing is essentially telling a *story* built from logical thought, decisions, and data outcomes. In the initial stages of writing, outline or sketch the series of figures and tables that will contain the data that tell your story. Further, to

convince your audience that your work is sound, you must do more than tell them *what* you did; you must also tell them *why* you made the scientific choices you made: Why this type of experiment? Why this particular computational model? Why these assumptions? Take the reader through the logical decision-making process that you followed. Show how these decisions fit into a larger research plan.

5. Good scientific writing, while thorough, is concise and expresses concepts in the minimum space necessary. Each paragraph and sentence should contain *substantial new information*, rather than largely repeating concepts already expressed and established. The exception to this would be in a larger document in which some repetition is helpful at widely separated locations to remind the audience of important concepts. For instance, a conclusion section would acceptably contain repetition of previous ideas.

6. You cannot presume an audience will simply trust your conjectures or opinions. Whenever you present an idea or finding that is not generally understood or accepted in the scientific community, or is likely to be questioned by your audience, you must *support it* by mathematics, logic, or data or provide a citation that does the same. If the matter is not central to your work, a citation is generally preferred (to keep discussion concise). If you cannot support a questionable statement, then you must qualify or narrow the statement so that it is supportable, or eliminate it entirely.

7. Relatedly, you should express a proper amount of scientific humility, objectivity, and self-skepticism—this will give your work more credibility and make embarrassing mistakes less likely. Outside of pure math, virtually nothing in science can be absolutely proven or is 100% certain. Instead, there are things that are more or less likely to be true or accurate, based on the available evidence and assumptions used. Avoid strong categorical and superlative terms like *always*, *never, all, none, wrong, right, best, worst,* and *most*. Instead express scientific judgments in more measured and narrow terms, and when possible quantify your certainty using appropriate statistical analysis (see rule 30). Compare the following expressions:

   - "this result is correct" vs. "this result is accurate at a 95% confidence level"
   - "this is the best explanation" vs. "this is a more satisfying explanation"
   - "the data prove that" vs. "the data suggest that"
   - "we are sure that" vs. "we have greater confidence that"
   - "this measurement has never been done before" vs. "this measurement has not to our knowledge been done before"

8. You must provide a bibliographic citation for any unique ideas, text (even if you change a few words around), images, or data you obtain from another source. While it is less common in scientific writing (paraphrasing is more often used), if you use the exact words of another, they must be enclosed in quote marks or otherwise set apart. Not to give due credit to another's work is called *plagiarism* and can lead

to extremely bad professional consequences. Additionally, citing others' work is a form of generosity and gratitude.

9. Copyright is a different concept than plagiarism, though they can overlap in some situations. Copyright law requires that you get permission from the copyright holder of an original work before using a significant portion of that work subsequently, even if you provide a citation to the original or it was created by you. Sometimes there is legal uncertainty about how much included material would be considered copyright infringement, but in most instances there are generally accepted practices. For instance, if you want to use a copy of a figure from a scientific paper in your own public document, you must contact the publisher of the journal for permission to do this, which permission they generally give. Many scientific journals now give an author or co-author automatic permission to republish their original content in a later publication (such as a thesis or dissertation), but you should still check the copyright policy of the publisher.

Organization and Scientific Grammar

10. Paragraphs that lack a proper topic sentence, are too long, and that jam multiple uncohesive ideas together are one of the most common problems of beginning scientific writers. Every time you write a paragraph, ask yourself: "What is the *single main idea* of this paragraph?" That idea should come out in the first or second sentence, which is known as the topic sentence. Every following sentence should provide information that closely relates to the topic sentence. If I were to read only the topic sentences for the whole document, I should get every important idea, if not the details. Again, do not switch ideas in the middle of the paragraph even if the second idea seems somewhat related; instead start a new paragraph. Paragraphs of even one sentence are allowed, when occasionally necessary.

11. There should be a new section or subsection at least every two pages of text.

    (a) Each needs its own heading and introductory paragraph. This introduction summarizes the main ideas and logically connects the present topic to the prior topic. Frequent roadmaps like this keep you as a writer focused and keep the reader from getting lost.

    (b) Generally you should not create a section or subsection unless there are at least two parallel elements (i.e. sections on the same level).

12. Technical writing requires a careful choice of words and punctuation to convey meaning clearly and concisely. With each sentence or other construction, ask yourself: "Is there any way for a reader to misinterpret the intended meaning of this statement?" Here are some examples of problem areas:

    (a) Confusion can arise if you use an anaphor (e.g. *it* or *this* or *that*) to refer to an antecedent (a previous idea, object, or person) and it is not clear to which thing or things you are referring because there are multiple possibilities.

Example: change "the above discussion demonstrates that it is not true" to "the above discussion demonstrates that the adiabatic assumption is not true."

(b) Avoid use of the slash character (/) as a conjunction: "liquid/gas mixture" could unclearly mean "liquid and gas" or "liquid or gas." Similarly, the construction "and/or" may not be entirely clear. Instead of saying "liquid and/or gas" say "liquid, gas, or both."

(c) Compound adjectives abound in technical writing. A compound adjective is a group of words that operate together to modify a noun. Such often require hyphens to eliminate possible misunderstanding. As an example of the importance of hyphens in compound adjectives, note the difference in meaning in the following phrases: "man eating dog" vs. "man-eating dog," "twenty two minute delays" vs. "twenty two-minute delays," "out of the box solution" vs. "out-of-the-box solution," and "more common species" vs. "more-common species." The exception to the use of hyphenation is if an adjective and noun are so frequently used together that the meaning is clear. For instance, "molecular dynamics simulation" is acceptable without a hyphen because "molecular dynamics" is a common term in theoretical chemistry.

13. Lists have some particular requirements:

    (a) Make your lists complete: avoid use of "etc." in lists because it requires logical extrapolation and therefore your intended meaning may be unclear.

    (b) Whenever you form a list, each item should be grammatically parallel with the others. For instance, if item 1 starts with a verb, then so should all remaining items.

    (c) Sometimes you need to assign an attribute to each item in a list or to compare two lists item-by-item. To make the assignment or comparison more concise, one can use the terms *respective* or *respectively*. For example, "Equations 1 and 2 are respectively mass and energy balances," or "the first three results were positive, negative, and negative, respectively," or "respective model and experimental data are given in the columns of Table 2."

    (d) Never make a bulleted list with only one bullet (analogous to rule 11b above).

14. When discussing upcoming sections, figures, and equations in the present document, or discussing enduring scientific principles, use verbs in the *present* tense. When discussing human actions, such as experimental data collection or the work of others, you may use the appropriate *past* or *future* tense. Examples:

    - Section 2 <u>describes</u> prior experimental results <u>collected</u> by Turner et al.
    - Kinetic energy <u>is given</u> by the following formula.
    - The data set <u>measured</u> using this instrument <u>is presented</u> in Table 2.
    - We <u>anticipate</u> this technology <u>will be used</u> to improve performance.

15. Consistently apply formatting and capitalization conventions, especially for the journal to which you are submitting. Some particulars:

    (a) For the headings of sections and subsections you should either follow the convention for capitalization of a *title* (mostly capitalized) or to capitalize the heading like a *sentence* (only the first word and any proper nouns).

    (b) Be consistent in how you capitalize and abbreviate a cross reference to a numbered section, equation, or figure. For instance, choose one of the following formats to refer to figures: "Figure 2," "Fig. 2," "figure 2," or "fig. 2."

    (c) Either italicize or don't italicize all Latin expressions (e.g., i.e., et al., vs., etc.)

    (d) Capitalize so-called proper adjectives derived from names: Ohmic, Faradaic, Fickian, Darwinian.

    (e) Chemical *symbols* are always capitalized (e.g. Ar, H_2O_2, Li-ion battery), in contrast with expanded chemical *names* (e.g. argon, hydrogen peroxide, lithium-ion battery). Chemical names are only capitalized if they are the first word of a sentence or are part of a title or heading, according to convention.

16. When referring to others' work by name, put the numerical citation immediately following the name or at the end of the sentence. If there are two authors list both last names: "Jones and Wu [3] developed a method that..." If there are more than two authors of a paper, explicitly list the first author only: "Srinivasan et al. [4] state that..." If there are multiple publications from one research group, you may refer to the senior author or leader of the work: "Newman and coworkers [5–7] developed a series of models..."

17. The first time an acronym is used, it must be defined by using parentheses, e.g. "focused ion beam (FIB)." If your reader at some later point is likely to have forgotten an unfamiliar acronym (particularly for long documents), then define it again as a courtesy. Do not use inappropriate capitalization in your definition: "Focused Ion Beam (FIB)" is wrong, whereas "Federal Bureau of Investigation (FBI)" is correctly capitalized.

18. Some writers use an apostrophe to make a plural out of an acronym, abbreviation, or symbol. However, avoid this where possible: "two SEMs" is preferable to "two SEM's" and "three τ values" is preferable to "three τ's."

19. Some writers insist that personal pronouns and possessives (e.g. *I, we*, *my*, *our*) not be used in technical writing. This is an old tradition that is losing its hold. As appropriate you may occasionally use personal pronouns in describing your research and results, in order to avoid awkward passive constructions. On the other hand, do not overdo the use of personal pronouns: the science, not the people doing the science, should be the main focus of the writing.

20. The choice of article (*the*, *a*, or neither) on a noun depends on whether it has been introduced yet or is familiar to your audience as a specific instance. In

other words, you could initially say, "this work requires *a* new type of conductivity experiment." After this experiment has then been introduced you refer to it as "*the* conductivity experiment." Use no article on a noun when you are discussing a concept abstractly or collectively: "diffusivity is a measure of mass transport by local molecular motion" vs. "*the* diffusivity of helium is high because of its small molecular mass," or "batteries are electrochemical devices" vs. "*the* batteries made with this additive function better."

21. Whether to use *a* or *an* as the indefinite article depends on how something sounds when spoken aloud (*"an* SPH model" not *"a* SPH model"; "*a* unique process" not "*an* unique process").

22. Use the words *data*, *phenomena*, and *species* properly. *Data* is the plural of *datum*. Thus, "the data show" is correct and "the data shows" is not correct (though you could say "the data set shows"). Similarly, *phenomena* is the plural of *phenomenon*. *Species* is the singular of *species*. Thus, "this species is the most prevalent" is correct.

Figures and Tables

23. Tables are used for data sets to enable the reader to obtain precise values for their own use, or where there are data for only a small number of experimental conditions. For example, a table of parameter values used in a model is helpful. Otherwise figures are preferred for presenting data because of the brain's ability to rapidly assimilate information and detect trends when data are presented graphically or visually. Avoid putting the same set of data in both tabular and graphical form.

24. A figure in the form of a line drawing or schematic is necessary to describe an important experimental apparatus or model geometry that is unfamiliar to your audience. A conceptual diagram explaining the process steps, experimental design, or workflow of your project may similarly be helpful.

25. Optical or electron microscope images should have a scale bar overlaid on the image.

26. Generally speaking, in your plots represent experimental data with discrete points or symbols; represent fits to the data or theoretical relationships with lines or curves. If you use a series of graphs to represent related data, use a consistent system of symbols and colors to aid the reader. Use colors and patterns that enable the plot to be interpretable when printed in black and white, not just color (also consider using "Color Universal Design" in your choice of colors). If you use Microsoft Excel to generate plots, know that the default settings do not make for nice graphs in formal documents—you will need to adjust formatting so they look better.

27. Each table and figure should be numbered, and have a caption that contains a title and in many cases an additional description that allows the table or figure to "stand on its own" without other supporting text. Even so, each table and figure should also be referenced and discussed in the main body text. Other rules about captions:

    (a) Table captions go above the table; figure captions go below the figure. Additionally, tables may have labeled footnotes below the table.

    (b) The title of the caption (i.e. the first statement) is not a complete sentence though it does end in a period. If subsequent description is needed, formulate as one or more sentences following the title.

    (c) Multi-part figures (a, b, c, and so on) have a single caption that describes the parts in sequence. For example, "Plot of results for (a) prior model and (b) current model."

    (d) The figure caption should define explicitly any symbols or lines in the plot, if a symbol key is not present in the plot itself.

    (e) Mention any special conventions in presenting the data. For example, you might use captions that include descriptions like "lines between experimental points are a guide to the eye," or "overlapping points have been offset horizontally for clarity," or "formatting is identical to that used in Fig. 2," or "error bars represent 95% confidence intervals," or "not drawn to scale."

28. Make sure that figures display with sufficient clarity when displayed on a computer screen or printed. On the other hand, make sure that compressed forms of figures are used so that electronic file size is not too large. These constraints can be met by using appropriate file formats for producing and saving figures: *vector* format for line drawings and graphs (.eps, .svg, .wmf) and *raster* format for photos and images (.jpg, .tif). For convenience you may choose to use a .png or .gif (raster) format for a line drawing or graph if the figure has a limited number of colors. When using raster formats, make sure that dpi (dots per inch) is at least 300 for sufficient resolution.

Quantities and Equations

29. Format quantities appropriately: $h = .221$ is wrong, $h = 0.221\,\mathrm{W/(m^2K)}$ is right; $k = 1.2\text{E-}3\frac{W}{mK}$ is wrong, $k = 1.2 \cdot 10^{-3}\,\mathrm{W/(m \cdot K)}$ is right. Notice that numbers and corresponding physical units are not italicized. If possible place a "hard" or non-breaking space (ctrl+space) between the number and its unit—this prevents them from being placed on different lines of text.

30. Confidence intervals are typically presented in the form *best estimate* ± *margin of error*. Round off your margin of error to one or two significant digits: generally if the leading digit is a smaller number (1 or 2) then use two significant digits and otherwise use one significant digit. You should then round off your best estimate (typically a sample mean) so its precision matches the precision of your margin of

error. For example, $t = 131.773 \pm 2.4329\,\mathrm{s}$ becomes $t = 131.8 \pm 2.4\,\mathrm{s}$. Notice how the revised best estimate and margin of error don't have the same number of significant digits as each other, but they do have the same level of precision (i.e. decimal place of least-significant digit), and that the sample mean was appropriately rounded off to that level of precision.

31. All symbols or variables should be italicized with the following noted exceptions.
    (a) Greek-letter variables and named dimensionless numbers (e.g. Re, Pr, Nu) should not be italicized.
    (b) Chemical formulas and common mathematical functions should not be italicized.
    (c) Vectors and matrix variables should be in bold font, unless one is referring to an element: v_i is scalar element i of vector $\mathbf{v}$.
    (d) Descriptive subscripts and superscripts that contain multiple letters that form a word or an abbreviation should not be italicized (e.g. k_i^{eff}, $x_{a,b}^{\mathrm{max}}$, t_{avg}). Numerical subscripts or superscripts should not be italicized (e.g. t_0, $g^{(2)}$)

32. When a symbol is used in a title or heading or sentence, it must remain in the same case (capitalized or not capitalized) as it was originally defined. If you start a sentence with a symbol that is not capitalized, it will look odd and so avoid this when possible.

33. Whenever an equation is given, all symbols or variables contained in it should be defined if they have not been defined previously. If you introduce a symbol not familiar or intuitive to your audience, and it is not used for some time in your document, as a courtesy redefine it on the second instance of use. Strive to choose symbols that will be familiar to your audience, e.g. k_B for Boltzmann's constant, σ for stress, q for heat flux.

34. In addition, for a long and math-heavy document, as a courtesy you should include a *List of Symbols* or *Nomenclature* section, a table that summarizes symbols and associated superscripts and subscripts for the entire document.

35. An important equation should occupy its own line and be numbered. It is not necessary to explicitly punctuate the equation with a comma or period, but it is still considered part of a sentence. Example: "The ideal gas law is

$$PV = nRT \tag{1}$$

where P is absolute pressure, V is volume, n is number of moles, R is the universal gas constant, and T is absolute temperature." Use the same equation editor when defining and using variables in the text as is used for the full equation, so that they appear with the same font.

36. Do not reference the equation number in the same sentence that contains the equation. Also, do not indent the line of text that follows the equation, unless it is part of a different paragraph. For instance, don't do this: "The ideal gas law is given by Equation 2:

$$PV = nRT \tag{2}$$

where P is absolute pressure, V is volume, n is number of moles, R is the universal gas constant, and T is absolute temperature."

37. A less-important or smaller equation can be given as part of a line of text (a so-called in-line equation) and must be formatted so that it is not too tall or the font too small. For instance, the in-line expression $\dfrac{N^2}{3}$ could be better formatted as $\frac{1}{3}N^2$ or $N^2/3$, and $\frac{d\left(1-\frac{\beta}{t}\right)}{dt}$ is better formatted as $\frac{d}{dt}\left(1-\frac{\beta}{t}\right)$ or $\frac{d}{dt}(1-\beta/t)$.

38. Do not format your equations how they would look in computer code. Instead learn how to use an equation editor so that parentheses, integrals, fractions, and other elements are sized appropriately. For complicated expressions nest parentheses inside of brackets inside of braces, i.e. $\{[(\dots)]\}$. Compare the two ways of formatting the following equation:

$$(\textstyle\int((n-1)/(n+1)+n^3)dn)^2$$

vs.

$$\left[\int\left(\frac{n-1}{n+1}+n^3\right)dn\right]^2 \tag{3}$$

Similarly, with the exception of named dimensionless numbers (e.g. Re, Pr) and well-accepted combinations (e.g. δt, $\Delta\phi$), avoid using symbol names with multiple letters, like you might do in computer code (e.g. C_{test} is better than $Ctest$).