Pet peeves, bêtes noires and just plain bad writing in journal papers, theses and dissertations[©]

by
Ben R. Hodges, Ph.D.

Department of Civil, Architectural and Environmental Engineering
The University of Texas at Austin

4th edition (in preparation)
April 19, 2006

Abstract

Writing is typically on par with getting a root canal for science and engineering students. The world of creative writing and English grammar rarely held their interest, so their writing style (or lack thereof) develops by imitating journal articles and textbooks. Unfortunately, they also learn to write dissertations and theses by following the rather dubious examples of prior students who "got by." Furthermore, I am continually appalled by the lack of rigor in logic and argumentation in many first draft manuscripts (culminating in "it is important" merely because I assert it so). As my students' (and my own) mistakes are likely universal, the present work is an attempt to provide some prophylaxis or (if nothing else) a means of reducing my commenting workload[*].

---

[*] DISCLAIMER: I have no authority or legitimacy as a critic of English grammar, syntax or usage. My main focus is on logic and clarity of communication. I do not claim to remember (or adhere to) all the arcane standards of written English. I do not recall the difference between a gerund and a participle (no doubt leaving the latter dangling). I have boldly split infinitives, and don't really pay attention to the preposition that I end a sentence with. It follows (see the logic?) that you will no doubt find multiple English sins in the following pages. Furthermore, this paper contains merely my own prejudices and quirks, which I feel free to violate at my whim – as Emerson once wrote: "A foolish consistency is the hobgoblin of little minds, adored by little statesmen and philosophers and divines." The diligent student should really invest some time with Strunk and White and a class in elementary logic and argumentation.

Table of Contents

A. The relationship between your writing and your work

A.1. Writing as science, not hand-waving:

Don't separate science and writing – writing is an integral and critical piece of the scientific process. If you don't write it down, it isn't science. Remember how annoyed you get at professors who stand in front of the class and wave their hands about ideas rather than writing them down on the board? For your writing to be good science, each statement in your work should fit into one of several categories. Every sentence is typically about something that...

a. ...someone else has definitively shown, which requires a citation;

b. ...someone else has argued to be true based on evidence, that may be either supported or refuted by yours or other work, which again requires a citation;

c. ...you are showing, which requires a clear linkage to evidence presented in tables, graphs and equations;

d. ...you are claiming to be demonstrably true based on your notes and/or experiments, but is so trivial that it doesn't need to be shown, which requires "(not shown)" after the statement – note that in general a dissertation should show everything that you have demonstrated to be true, if it is trivial it can go in an appendix;

e. ...you are arguing to be true as a consequence of something that someone else has shown or something that you are showing; note that it should be clear that this is your argument, or that it supports or refutes someone's argument – whenever you are showing by argument, you must make a clear and logical case that follows from some evidence;

f. ...you hypothesize may be true and that you have some evidence for, but may not be able to definitely demonstrate;

g. ...is pure speculation that provides an explanation for something observed, but cannot be considered to be irrefutable evidence, i.e. you may have evidence that supports your speculation, but other speculative explanations could readily be advanced that would also be supported by the same evidence. You need to be clear in letting the reader know how your evidence fits into others evidence, argument and speculation. Watch out for places where you assert something to be true without providing some form of evidence or argument to back it up.

A.2. Big picture first, details second

Give the reader an understanding of the big picture before you get to the details. That is, don't start a section by building up descriptions of minutiae and then at the end of the section tell the readers what they should have figured out. Start an explanation by stating what you are going to show and the basic means of getting there, then give the details that back up what you want to show. Imagine if someone analyzed an elephant by giving you descriptions provided by the proverbial blind men, and then told you the result is an elephant. This approach is fine if the descriptions are short and you can keep them all in mind at the same time, but it is better to say that the creature is an elephant, a large four-legged beastie with a long trunk, leathery skin, a small tail and thick legs. Afterwards you provide the details of the toenails, the tusks, the big ears, etc. This concept is critical in doing mathematical derivations – just because you got to the end by struggling through a complicated path doesn't mean your reader should also suffer. Tell them where you're going then show them how to get there.

A.3. Don't slip into a chronology

Avoid phrases such as *this,... then that*, or *after that*, or *once this is done*...etc. Think of your research as an existing whole – your objective is to explain this whole, not step by step process you took to get there. The only time a chronology is acceptable is when you are describing a precise series of steps that must be followed (e.g. a lab experiment). When you write a chronology there are two critical problems: first, you bore the reader to death, and second, you end up sounding whiny – "look at all the work I had to do...." The amount of time you took for a particular step of your work has no bearing on its scientific importance. Often we spend most of our time on things that later turn out to be unimportant.

A.4. Research the literature

A literature review needs to be both deep (in relevant works), and wide (in works which touch on the topic). Relatively few areas are so new that less than 30 or 40 relevant references can be found. For proposals on new research, I routinely include 25 or 30 citations to the literature to demonstrate the connections between my work and other work.

A.5. "The figure shows..."

... is an anthropomorphization (see B.6), that invites the reader to see data exactly as you do. Make sure your figure clearly reflects your claim rather than your bias or your particular expertise. Too many figures only "show" the claimed result when the reader has exactly the same knowledge and experience as the writer. Many figures "show" the result only when you squint real hard, hold it at a distance and turn it slightly sideways (I achieve the same effect by removing my glasses). Figures should: 1) illustrate concepts that are difficult to explain in the text, and 2) present data that supports your interpretation. The text accompanying the figure should help the reader get to the meaning of the figure.

A.6. "Comparing figure 1 and 2 indicates..."

Inviting readers to compare figures is inherently a qualitative exercise and is fraught with peril. If you say they are similar, be ready for a reviewer to point out all the dissimilarities. If you say they are different, make sure the differences are large and obvious. Generally, I prefer to invite comparisons only for differences – and then principally when it is difficult to make a meaningful quantitative analysis. If data is sufficiently similar, you should be able to make a quantitative computation of the difference (e.g. the RMS, or some error norm). Additionally, try not to have figures show things by forcing the reader to compare to figures from different pages; figures to be compared should be in separate frames of a single figure.

A.7. The importance of being quantitative

Avoid words like ‘important’, ‘significant’, and ‘critical’. They generally indicate that you don’t have enough information to quantitatively discuss the subject. Furthermore, if you must use one of these words, e.g. “important,” it should be as a *qualitative* assessment, so something must be “unimportant”. Thus, you may say that A is important compared to B, but it is nonsense (from a scientific point of view) to simply assert that A is important (see also B.15, **Primary implies a secondary**).

A.8. What is your contribution?

Make sure you clearly separate what you have done from what others have done. This problem is often related to “**What is the it?**” in your sentences (see B.13) and the use of third person perspective. Beyond the grammatical aspects, you need to make sure that reader does not think either: 1) all the work you describe was done by others, so you made no contribution; or 2) you are claiming work that was actually done by others. In the first case, your thesis fails because you haven’t made a contribution, in the second case it fails because you have plagiarized. Thus, make sure to emphasize and distinguish your contribution from prior work.

A.9. Declaring facts is annoying

Phrases like, “the fact that...” or “in fact,” or “it is a known fact” are annoying as they are transitions where you are asserting the truth of something and claiming that you don’t require evidence. Sorry, I want evidence. To me, a fact is something that someone has definitively observed. That is, it is indeed a fact that Osbourne Reynolds observed dye streaks that remained essentially uniform or quickly mixed, depending on the speed of the flow. If you really want to be pedantic, it is a fact only that he reported these observations (while I do not have personal knowledge that extends beyond his paper, I do tend to give credence to the literature and considered the reported observations to be facts). However, the explanations given for observations cannot be considered facts. It is not a “fact” that Osbourne Reynolds observed the transition from laminar to turbulence. The words laminar and turbulent are products of continuum mechanics, which is a *theory* that describes everyday existence, but is (in fact) wrong. Best to avoid the use of the word “fact” unless it is critical to point out the difference between something observed (a fact) and an explanation (theory or hypothesis).

B. Tips on style and grammar

B.1. And a good sentence has only one "and"

This rule improves the logic and clarity of your work. When you have more than one "and" in a sentence, it is generally unclear what items you are grouping together. Often, two "and" implies two different groupings, but this won't be clear. Any sentence with more than one "and" should be reworked to form multiple sentences; you probably have more than one distinct thought you are trying to say. I learned this from a sign on the classroom wall of my son's first grade teacher and I have found the idea useful and it is often abused and its violation makes for really long sentences and we could go on and on.

B.2. Repetitive explanations

Repetition may be part of your writing process, but some culling will help. When I write, I often write a thought, and then I follow the thought with another way of explaining it. That is, I say something then I say it again in a different way. Another way to look at this is that I keep saying the same thing until I find a way to say it that makes sense. Moreover, I sometimes repeat myself. Enough already! Hopefully you get the point: a sentence that is well-written shouldn't need amplification or further explanation. Naturally, there are exceptions – when an idea is difficult, it may be useful to present it in two different ways or give an example (see B.5, i.e. and e.g.). However, this is a rule that I don't try to enforce when I'm writing – I go ahead and repeat myself because often the third or fourth sentence is the best one, and the others end up deleted later. If you write like I do, make sure you carefully edit your work for repetition before passing it on for review!

B.3. Repetitive transitions are repetitive

Watch for overusing transitions such as "therefore" or "since." If every one of your sentences takes the form "Since A is true, B is also true" or "B happens since A occurred," your writing will become tedious. Also, recognize that these transitions imply a logical causality between what comes before and what comes after. Make sure that "A therefore B" makes logical sense.

B.4. Repetitive sentence beginnings are boring

If two sentences in a row (or two headers!) begin similarly, the reader gets bored and easily loses focus. For example: "This section shows that A=B and C=D. This section is therefore really boring..." Note that it is OK (indeed desirable), to repeat key words that you want to stick with the reader – just don't keep repeating the same filler and transitional words.

B.5. i.e. and e.g.

Check out the movie *Get Shorty* for a memorable discussion. The first is the equivalent of saying "that is" while the second is the equivalent of "for example." These are useful (if not overused) to help tell the reader that the next sentence is not an additional thought, but is an amplification of the previous thought. Too many i.e. and e.g.'s sprinkled throughout your writing indicate you are probably writing sentence that are overly complex (see also B.2, Repetitive explanationsabove).

B.6. Don't anthropomorphize

It's a good word, look it up. As an example, people can assume things, inanimate objects (such as research) cannot. From your data, you might infer a conclusion, but the data themselves cannot do so. Unfortunately, we usually do allow data to imply things (which is not quite as annoying as growing the economy, but it is getting close). The most common anthropomorphization is probably the least objectionable: "The figure shows..." is clearly an anthropomorphization, but can be used when the more correct alternatives seem too wordy. Alternatives are "As shown in figure..." or "As can be seen in...". All these phrases have other implications discussed in **"The figure shows..."** (see A.5).

B.7. **Watch your ass**umptions

Be careful with words like "assume". Save the word "assume" for when you need it in its true technical sense – i.e. when you don't have any data and you have to make an assumption to solve a problem. In which case, you need to argue reasons why the assumption is valid. Don't use "assume" when you really are saying that you are limiting your focus to a subset of cases or are making an approximation. You cannot assume the world is flat and that the sun revolves around the earth – however, if you limit your solution space to scales that are small compared to the earth's rotation and curvature, you can neglect the effects of curvature and rotation and use Newton's equations of motion for simple Euclidean geometry. This has the same effect as assuming the world is flat and motionless, but doesn't make people think you're stupid or your work trivial. Often the word "approximate" is much better than "assume" – i.e. Newton's laws are a good approximation for continuum behavior, we don't need to assume that they apply.

B.8. In order to make your work concise

Delete "in order" throughout your paper. The statement "in order to do something" can almost always be replaced by "to do something".

B.9. Limit new acronyms

It can be annoying to readers if you create too many new acronyms. In the extreme, this can lead to sentences such as "When analyzing the results of an RBC test with the NFE methodology show that SFE can be FRE with the KDL under limited circumstances." The key to deciding if you need to use an abbreviation is 1) if you will use it often, i.e. every page or so; 2) it is critical and central to the work – i.e. don't coin new jargon for peripheral issues, and 3) the words you are abbreviating are long and annoying.

B.10. This is indefinite.

Whenever you use "this" or "these' it should be followed with a noun, not a verb – e.g. 'this equation', 'this figure', 'these authors' is correct. Avoid "this is", "this shows" "this means," or similar constructions, which leave the reader wondering "this what?"

B.11. Be tense about tense

Work that other researchers have completed should be referred to in past tense. The work you are presenting is present tense, unless you are specifically describing something that happened on a particular day during the course of the research. It is relatively easy to keep your research in the present tense if you "Don't slip into a chronology" (see A.3). One way to think about this is to use past tense when discussing something that was observed or completed in the field or laboratory, but use present tense for what you are expecting the reader to see in your graphs, figures and equations; i.e. "The instrument *was deployed* such that..." is proper past tense, whereas "Equation (15) and figure (4) *illustrate* the connection between..." is proper present tense.

B.12. You must limit your imperatives

Declaring something "must" occur or "must" be done is setting up an imperative or an absolute. As a reader, I often get sidetracked by "must" as I start thinking about other options that prove the author wrong so that the statement should really only be "may." You should be willing to bet your B.S., M.S. or Ph.D. degree on the absolute truth of any statement that includes "must" or any similar imperative.

B.13. What is the it?

The word "it" can be troublesome. In particular, I've never really liked (although I admit to having used) the common scientific jargon that "It is shown that...." Who or what is the "it"? The alternative for "it" is usually a switch to the first person, "We show that..." which may not be appropriate (depending on the venue). Personally, I don't mind the first person frame, as it helps separate what you have done from what others have done. "It is shown" is sometimes used for what someone else has done as well as for what you have done, so you tend to be downplaying your accomplishments by saying "it is shown" (see also A.8 **What is your contribution?**). We may be dispassionate observers (third person) for our work, but we are the ones doing the demonstrating and showing (not some mythical scientific 'it'), so we might as well take credit!

B.14. Technical jargon – the good, the bad, and the ugly

Every discipline has its jargon. The 'good' is that jargon allows you to say something in a very precise way that will be clear to those familiar with the jargon. Thus, when I say "frequency" the jargon-knowledgeable reader should understand that I am referring to a system property defined as 1/T, where T is some time measure. The 'bad' is that much jargon is adapted from common words with ancillary meanings. Thus "frequency" can be used in a non-jargon (hand-waving) sense, such as "The frequency with which we see this result leads us to conclude...." which really isn't about the 1/T value, but is really about the qualitative dominance of the particular result in the data set. The "ugly" is when the jargon gets so thick that the meaning of sentences cannot be adequately deciphered. What should you do? Firstly, don't avoid jargon, since it can considerably simplify your writing by making it more concise (i.e. you don't repeat things the knowledgeable reader already knows). Secondly, don't overuse jargon (i.e. don't use it for its own sake, as obfuscation; use it to make your thoughts clear). Thirdly, only use "jargon" words in their precise jargon meaning. To continue the example above, if you use “frequency” in both the jargon and non-jargon meanings, you effectively make yourself more confusing. Some jargon words to watch for are: frequency, deviation, accuracy, error, precision, system.

B.15. Primary implies a secondary

If something is primary, then something else must be secondary. The secondary items should be explicitly enumerated, unless it is easy for the reader to infer the secondary item (or items). Otherwise, you are simply using “primary” as a hand waving adjective to assert importance (see also A.7 **The importance of being quantitative**).

B.16. Between ‘a’ and ‘the’

English articles ‘a’ and ‘the’ are subtle signals to the reader as to whether the following item is general or specific. For example, the simple phrase “***A*** velocity comparison method is used to...” is significantly different from the phrase “***The*** velocity comparison method is used to...” In the former case, the reader is invited to consider the words “velocity comparison method” as a general description of the approach taken. In the latter case, the word “the” implies a specific method, known as the “velocity comparison method” has been applied, so the reader needs some explanation of the method. For example, you might write that “The velocity comparison method (see section 4.2) is used to...”

B.17. Noun-ification

Noun-ification is a made-up word[†] describing use of verbs as nouns in prepositional phrases (i.e. phrases beginning with prepositions such as of, with, for, to, in). *This usage results in confusion of the reader.* The previous seems like an OK sentence – at least grammatically. However, “confusion” is the noun form of a perfectly good verb that provides a clearer sentence: *The previous usage confuses the reader*. Both sentences are technically correct, but the latter is clearer[‡]. English speakers often formulate their speech and thoughts with lots of prepositional phrases, which creates passive sentences (no action verbs). Indeed, when I write informally (as in this document), I tend towards passive writing. My hypothesis is that prepositions give us more thinking time when we’re speaking, which carries over to our writing.   I don’t try to solve this problem when writing; editing while writing slows me down. Instead, I generally look for prepositional phrases that can be simplified to active verbs (activated?) *only* after I have finished getting the basic structure of a paragraph onto paper.

[†] Dare I claim that I made it up? There may be some truthiness involved (apologies to Stephen Colbert and the writers of *The Colbert Report*).

[‡] Thanks to Prof. Chris Rehmann of the Civil Engineering Department at Iowa State University for this example.

C. Formatting and layout

C.1. Reference figures now, not later

Don't put the figure reference at the end of a long discussion. Let the reader know at the start of a discussion that there is an accompanying figure. There is nothing more annoying than puzzling through a complex explanation and finding "as shown in Figure 2" at the end. This issue is absolutely critical when the figure ends up on the next page.

C.2. Define your acronyms once

And don't switch back and forth between the acronym and the full text. I can think of only two exceptions to this: 1) in a long dissertation it is OK to restate the definition of an acronym (especially a non-standard acronym) when it is first used in each chapter; 2) it is sometimes useful to use the full text rather than the acronym in a concluding paragraph to make it more readable to people (like me) who start reading with the conclusions.

C.3. Learn to use the Word Style functions

Do this now, and save yourself many headaches later. Instead of formatting a paragraph or word, you should develop and use a style that is consistent through the text. You should define a 'Body' or 'normal' style for all paragraphs that includes the appropriate indentation and line spacing etc. You should be able to change the line spacing through the entire document by simply changing one style setting. Look at the Format – Styles and Formatting menu to see what styles you are using. Also, use the Word features for headings and numbering based on headings. These tools allow automatic creation of table of contents etc. All your equation number references should be automatic cross-references that change when the equation number changes

C.4. Equation symbol definitions

Where possible, use standard symbol definitions for your subject (e.g. in fluid mechanics ρ is density, 'u' is the x-direction velocity). The first time a particular symbol is used in an equation, it should be explained below the equation. Such definitions may not be necessary in a short report that uses standard terms defined in a table of nomenclature at the start of the report. However, in long reports, it can break the reader's concentration to continually refer to the table of nomenclature for obscure symbols. When there is a single term on the left-hand side of an equation, it is sometimes useful to define that term in the sentence before the equation. Within a particular section of a thesis or dissertation, you do not need to keep defining terms that were defined in prior equations. However, you should use some judgment. In general, it is not a bad idea to provide the definitions of obscure symbols anew in each chapter of a dissertation or thesis, but it gets annoying when common ones are provided for every equation. On the other hand, if the last use of a previous non-standard symbol was 10 pages prior, then a repeated definition of an obscure symbol may not a bad idea if it keeps the reader in the flow of the text.

C.5. Equation consistency

Make sure you include a table of nomenclature in a thesis, dissertation or technical report. The table should have all the symbols used in your paper. No symbol should have multiple definitions. Symbols used in an equation must be consistent in typeface. That is '**A**' is not the same as 'A' or 'a' or A or A . However, this doesn't mean you should have 5 different 'A' definitions using slightly different typefaces. For journal papers, make sure you know what the editor requires – some want nomenclature tables, others do not.

C.6. Equation format

Take a look at equations in textbooks and use these as your patterns. In general, you will not see either '*' or 'x' used for multiplication, so don't write *a x b* where *ab* can be used. Modern word processing gives you the tools to make your equations clear, professional and unambiguous.

C.7. Equation parentheses

Always use the automatic sizing brackets and parenthesis in equation editor; i.e. use

$$\mathrm{a} = \mathrm{bx} + \mathrm{c}\left(\frac{\mathrm{z}^2}{2\alpha} - 4\right) \quad \textbf{\textit{not}} \quad \mathrm{a} = \mathrm{bx} + \mathrm{c}(\frac{\mathrm{z}^2}{2\alpha} - 4)$$

C.8. Subsection organization

C.8.a. Subsections get lonely

They just cannot stand alone. For example, we are presently in section C.8.a, which means that there must be a C.8.b. Otherwise, there is no logical justification for putting this paragraph in a subsection.

C.8.b. "Three shalt be the number thou shalt count, and the number of the counting shalt be three... Five is right out" [§]

Some people get a little carried away with organizing by section headers. Subsections are generally OK at about three levels. I can keep in mind that I'm in the second subpoint of the 7th idea in the 3rd section of this paper. This is still somewhat meaningful. However, once you start into the 4th level...

C.7.b.1 Sub-sub-subsections

Or into the 5th level...

[§] with apologies to Monty Python. For the more demented, the full quote (from *The Holy Grail*) is: "And the Lord spake, saying, 'First shalt thou take out the Holy Pin, then, shalt thou count to three, no more, no less. Three shalt be the number thou shalt count, and the number of the counting shalt be three. Four shalt thou not count, nor either count thou two, excepting that thou then proceed to three. Five is right out. Once the number three, being the third number, be reached, then lobbest thou thy Holy Hand Grenade of Antioch towards thou foe, who being naughty in my sight, shall snuff it.' "

C.7.b.2.1 Drowned sections

... you are creating something with about the same organization of the IRS tax code, and about the same readability. Anyone that can look at a set of five subsection labels and can accurately recall all the higher level subjects is obviously *not* an absent-minded professor (and is therefore probably not on your reading committee). Furthermore, if you need 5 levels, you've probably got problems in your argument structure. It may be that some subsections are irrelevant to your main point and can be relegated to an appendix.

C.8.c. Introductions are also subsections

You shouldn't have a "zeroth" paragraph. That is, consider the following bad example:

> **1. Section Title**
>
> Here is some introductory text explaining this section. blah blah blah and yada yada yada or anything else you want to say
>
> ***1.1 Subsection one title***
>
> Here is the start of the first subsection, again its more of the same blah blah blah and yada yada yada or anything else you want to say
>
> ***1.2 Subsection two title***
>
> Here is the start of the second subsection, again its more of the same blah blah blah and yada yada yada or anything else you want to say

The problem with the above is that if you wanted to provide a cross-reference to something in the introductory paragraph, what would you do? If you say Section 1, it refers to all of Section 1, not just to the introductory paragraph. So the better approach is

> **1. Section Title**
>
> ***1.1 Introduction***
>
> Here is some introductory text explaining this section. blah blah blah and yada yada yada or anything else you want to say
>
> ***1.2 Subsection two title***
>
> Here is the start of the first subsection of real stuff (as opposed to introductory stuff), again its more of the same blah blah blah and yada yada yada or anything else you want to say
>
> ***1.3 Subsection three title***
>
> Here is the start of the second subsection, again its more of the same blah blah blah and yada yada yada or anything else you want to say

C.9. Tables with numbers...

...belong in an appendix: Before you put in a table with numbers, ask yourself what use the reader will make of the numbers. If the numbers are necessary for archival purposes, then they belong in an appendix.

...or should have two significant digits: If the reader actually needs the numbers, they shouldn't have more than two (or three at a stretch) significant digits and the table shouldn't be more than a half-dozen entries. A table with 14 numbers at 4 or 5 significant digits is just too much information for anyone to comprehend in a simple manner.

...or better yet, should be a bar graph: If the reason you are putting in the numbers is to show their interrelationships, then they belong in a bar graph. It's easy to do in Excel, so there isn't a good reason not to.

C.10. Figures, labels, captions and in-text descriptions

Figures are critical to any paper, but can lead to more confusion if you don't provide good labels, captions and descriptions.

Labels: Labels are required for all lines, symbols and axes in a graph. Ideally, keep to a single font size throughout the graph. Make sure the font is readable! Check with the journal on what size standard figures should be so that you know whether the figure will be shrunk and the font reduced. All values that have dimensions should be noted; i.e. if your x-axis is a length in meters, it should be noted labeled something like: "L (m)".

Captions v. in-text descriptions: Figure captions should not repeat what is in the text, nor should they be so brief that the reader must search through the text to understand the figure. In general, a figure and its caption should provide enough information for the reader to understand the importance of the figure without having to read all the text. Subscribers to Playboy and scientists have a lot in common – they claim to read the articles, but mostly look at the pictures!

Font size: Make sure your figure has readable fonts. For a thesis or dissertation, the equivalent of 10 point is my minimum. Journal papers often can go as small as 8 pt. Keep in mind that when you shrink the figure to fit the page, you also shrink the font. I write a separate matlab script for each figure in a paper; if I need to change the font size or figure size it usually requires only changing a single line of code.