Writing Scientific Papers

Generalities

Compiled by Dustin Haines

Common Word Usage and Formatting Errors

"Specie" is not the singular of "species." "Specie" is currency in coin form. The singular of "species" is "species."

"Data" is a plural term. The singular of "data" is "datum." In other words, use "data were collected," not "data was collected.

Always italicize a scientific binomial species name, but only the genus is capitalized, like this: *Lemna gibba*, *Homo sapiens*, *Falco peregrinus*, *Upupa epops*.

Choppiness

Avoid repetitive use of short sentences when you can combine related thoughts into a somewhat longer sentence. We use the word "choppy" to describe this way of writing, so when you see "choppy" written on your paper, look for ways in which you can consolidate sentences.

To illustrate, here is an intentionally choppy paraphrase of the last paragraph:

Avoid repetitive use of sentences. You can combine related thoughts into longer sentences. I will use the word choppy to describe this. When you see the word choppy, consolidate your sentences.

The original version flows better. Note, however, that there is a reciprocal trap; very long, complex sentences are both tedious and difficult. If you want to see more information on these subjects, here are two online resources:

http://owl.english.purdue.edu/owl/resource/573/03/

http://www.harpercollege.edu/writ_ctr/comma9.htm

Wordiness

This is something that we all struggle with. In science writing, a lot of emphasis is placed on being succinct and to the point, especially through the elimination of words or phrases that do not add important meaning or content. In trying to tighten your writing, look particularly for prepositional phrases, which can often be eliminated without compromising communication. Below are several examples that we see repeatedly, with remedial guidance (in a prior draft, this last phrase read "…with some guidance on remedies.").

Based on My Previous Research, I Do Not Need To Say These Words

There are many things that we say conversationally that constitute unnecessary verbiage in scientific papers. Here are a few phrases that you should generally avoid. Use your word processing software to search for these phrases, and try to delete them without compromising clarity:

"has been shown to" – can either be replaced with "can", or deleted

"It is clear that…"

"Due to the fact that…"

Beginning to Start to Write About What You Did

This title may seem ridiculous, but we often seen the words "begin" and "start." Focus instead on the *action*, what you *did*. For example,

> "To begin collecting data, we started by removing the plants from the pots …"

Instead try this:

> "We removed the plants from the pots."

Use the search tool of your word processing software to locate the words "begin," "began," and "start;" see if you can modify your sentences.

First Doing This, Then Doing That, Then Doing This.

A related point: very often the most logical organization of methods and results is neither linear nor chronological nature. If activities and results are functionally dependent, then the presentation should emphasize this contingency. Quopoting the Methods section of the GRADING RUBRIC for this assignment,

> "Write this section with enough information so that a peer could replicate your experiment (but not in a "first do this, then do that" kind of way). Organize it logically (which will only be temporally if there is a logical necessity to the timing)."

Even when the functional dependence of events is sequential, there is usually little to be gained by using chronological language (words like "first," "then," "before," and "after"). As an example:

> "We put a coffee filter in each pot, filled them with perlite, added eight soybean seeds to each, and added one inch of perlite on top of the seeds." I did not use any of the temporal terms here, but I still got the necessary information across.

Sometimes, of course, the sequence of events is very important (e.g., "Thirty minutes *after* the acetylene injection, a 1 mL sample from each tube was injected into the gas chromatograph column.").

Reviewing or Citing Published Literature / Papers

When reviewing published papers, or making a citation within your report to make a specific point, many people use a "conversational style". Here are some examples:

> "I read an article that was about . . ."
> "In an article by Tilman et al. (1994), they describe the . . ."
> "One of the articles I read found a correlation between . . ."
> "Based on previous research . . ."

If you were *talking* with someone about the papers, this style would be fine, but in scientific writing, the important point is not that you read the paper, but what you wish to convey about the researchers' findings. In the third example, there is additional silliness: an *article* is an inanimate object; *it* did not find a correlation, the *authors* found the correlation. Here is a fixed version:

> "Tilman et al. (1996) found a positive correlation between N-fixation and phosphorus supply in prairie legumes."

You rarely need to give the title of an article or book within the body of your paper; restrict the full title to the Literature Cited/References section at the end.

When referring to published literature, you seldom need to summarize the goal of the research; you can simply report pertinent findings. On the other hand, if you are doing an in-depth review of a paper, then it *is* a good idea to portray the goals of the research; this can go either way, depending on the objective of your writing.

Here are a few resources online that address wordiness and how to eliminate it.

http://owl.english.purdue.edu/handouts/general/gl_concise.html

http://web.uvic.ca/~gkblank/wordiness.html

http://leo.stcloudstate.edu/style/wordiness.html

http://www.transaction.net/web/tutor/text/index.html

http://people.whitman.edu/~hashimiy/wordness.htm

Writing a "Methods" Section

Make sure to identify the statistical analyses you used to test your hypotheses. This will often be the last thing you describe in your Methods section. You do not need to say the same thing repeatedly when you use similar analyses for multiple hypotheses or different sets of data. You also don't need to incorporate all of the detail that the Statistical Techniques section of this lab manual uses; in this venue, we are explaining the statistics, how they work, and how they are used. You merely need to explain what you did. If what you did is unusual, tricky, or interesting, you may also want to explain why you did it.

In the *Lemna* paper, you need to say that you used regression analyses, and what kinds of regressions, and for what data/graphs (e.g., linear regression for the d*N*/*N*d*t* vs. *N* graph, 2^{nd}-order polynomial regression for the d*N*/d*t* vs. *N* graph) they were employed. You should specify the r-squared value from each analysis, and what it implies (hint: the lab manual section on statistical analyses talks about the r-squared value). For the N-fixation paper, you can simply say that you used one-way and/or two-way analyses of variance to test for differences in means between treatments or treatment combinations, and to test for interactions among treatments. You should also mention the confidence level you used for determining if your means are different; for our analyses, we agreed upon a 95% confidence level. In other words, p-values that are less than or equal to 0.05 ($p \leq 0.05$) indicate that your means are different.

If you refer to means and standard errors of the mean within the text of your paper (see Reporting Results, below), you should identify any acronyms or the format that you use in the methods. As an example, "Means and standard errors are represented by (mean ± 1 standard error)", or "One standard error (SE) is given for each mean."

Reporting Results

It is common practice to summarize all of your results and analyses in the "Results" section. For example, if you calculate *K* and *r* from a d*N*/*N*d*t* graph, you will want to present these values in the Results first. You can refer to them again in the "Discussion" if you need to, but you should never refer to analyses or data for the first time in the "Discussion."

You should refer to r-squared and p-values for all of the statistical tests that you perform within the text of the paper. People also generally refer to means and standard errors in the text of the results; if you wish to do so, there is some guidance below (and above under Methods).

Units of Measurement

All figures (graphs) and tables should indicate the units of measurement that the columns and axes are displayed in. If you display data anywhere in a paper, you must indicate the units of those data. For the *Lemna* experiment, the units of measurement for population size are in numbers of thalli, with the population size, or *N*, being the *variable* that is measured. If you have a graph that shows *N* versus time, you need to display, on both axes, what the units are. This holds as well for d*N*/d*t*, d*N*/*N*d*t*, N-fixation, biomass, etc. In published papers, people frequently display both the variable and units on a given axis, with the variable followed by the units in parentheses (e.g. "Population size (# of thalli)"). This way you know exactly what is being displayed.

Referring to Statistical Significance

Generally, "statistical significance" implies that some event is unlikely to have occurred by chance. Determining whether or not your results show statistical significance nearly always entails the calculation of a test statistic, which is then compared to some distribution of test statistics to determine a p-value. Different degrees of rigor (symbolized by the Greek symbol α) may be applied to reach a conclusion about whether a chance null hypothesis is rejected by one's experimental data. Vales of $\alpha = 0.05$ and $\alpha = 0.01$ are common standards. If your p-value is greater than $\alpha = 0.05$, there is a less than 95% chance that your data are different from a random result; thus under the $\alpha = 0.05$ standard of rigor, we would conclude that the result is not statistically significant.

A common misconception is that a high r-squared value from a regression analysis indicates statistical significance. While this may be true, you cannot technically say this without obtaining a p-value from an analysis of variance for the regression, and comparing this to your alpha level. It is not correct to say the following:

> "Since the r-squared value was 0.895, our data was significant."

The r-squared value, instead, simply indicates the proportion of variation that can be explained by the correlation between the two variables. A "high" r-squared value, like $r^2 = 0.895$, indicates that a large proportion of the variation in the data (here, 89.5%) can be explained by the relationship between the x and y variables (thus, there is not much unexplained variation). A "low" r-squared value (like 0.023) indicates that there is a great deal of unexplained variation (aka, "noise") in the data.

Wording and Formatting in the Results

Here are a few general guidelines for the Results section:

1) Always indicate r-squared and p-values in parentheses at the end of a sentence.

DO - The decomposition rate of maple leaves was significantly greater than that of pine needles ($p = 0.01$).

DON'T - Because $p = 0.01$ we concluded that the decomposition rate of maple leaves was significantly greater than that of pine needles.

2) Describe how the mean varies between treatments, but put the actual mean in parentheses.

DO - Habitat type had a significant effect on the number of adult worms caught ($p = 0.02$). A significantly greater number of worms were caught in deciduous forests with buckthorn (mean = 25; SE = 3.8) than in deciduous forest without buckthorn (mean = 13; SE = 2.5) or in coniferous forest (mean = 11; $p < 0.02$).

DON'T - Our ANOVA had a p-value of 0.02, so we concluded that habitat type significantly affected the number of worms caught. We found an average of 25 worms in deciduous forest with buckthorn, 13 worms in deciduous forest without buckthorn and 11 worms in coniferous forest. A Tukey test showed that there were significantly more worms in the deciduous forest with buckthorn.

3) Always refer to a table or figure with parentheses at the end of a sentence.

DO - Canopy openness was significantly greater in deciduous forest than in conifer forest ($p = 0.04$; Figure 2).

DON'T - Figure 2 shows that canopy openness was significantly greater in deciduous forest than in conifer forest.

4) In determining *r*, *K*, and the maximum sustainable yield from your Lemna graphs, it is best to *calculate* these values using the regression line equation that Excel determined for you, instead of estimating them visually from the graph. Nor should you use individual data points to determine these values. For example, the highest single data point on the far left side of the d*N*/*N*d*t* graph wll not give you a good estimate of *r*, because any non-zero *N* will impose a negative feedback. Use the y-intercept calculated from the regression line equation.

Biological editors recommend using the active, rather than the passive voice in scientific reports. Thus, instead of writing, "The plants were measured. . . ", you should write, "We measured the plants . . ."**Note that this convention may not pertain** in **other disciplines (e.g., chemistry).**

It is perfectly acceptable to use the first person. Third-person constructions ("The author did this, that, and the other thing.") are pompous historical artifacts.

Use the past tense when describing your experiment and its results. Use the present tense when discussing general properties of organisms.

Bear in mind Johnson's comment on Milton's Paradise Lost : "If its length be not virtue, it hath no other." Every word costs money in print, so keep it as short as possible consistent with a good job.

Finally, never apologize for lacking data; let the reader assume that you have done the best job possible.

Guidelines for Citation

By Leah Domine

The following guidelines for citation are modified from those recommended by the Ecological Society of America for authors submitting a manuscript to its journals. Many societies and journals recommend minor differences in format, but these guidelines incorporate the essentials.

Within-text citations and Reference Section: All references cited in the text must appear alphabetically in the References Section. Conversely, all references cited in the References must be cited within the text.

Reference Section

The references should be placed at the end of the manuscript, before the tables and figures. References should be alphabetized by the last name of the first author. Note that the second and subsequent lines of a citation are always indented (i.e. the first line is "hanging").

Example of format, content, and punctuation for an ARTICLE citation:

First-author last name, initials, second author initials, last name. Year article was published, Title of article. Journal Name, Volume number, pages of article.

Pollard, E. and K. H. Lakhani. 1987. The detection of density-dependence from a series of annual censuses. Ecology 68:2046-2055.

Example of format, content, and punctuation for a BOOK citation:

First-author last name, initials, second author initials, last name. Year article was published. Book Title (italics). Publisher.

Alstad, D. N. 2001. Basic *Populus* Models of Ecology. Prentice Hall, Upper Saddle River, NJ.

Example of format, content, and punctuation for a CHAPTER citation:

First-author last name, initials, second author initials, last name. Year the book was published, Title of chapter. In: Initials of Editor, Last Name of Editor [eds.], Book Title, Publisher.

Brown, J. H., D. W. Davidson, J. C. Munger and R. S. Inouye. 1986. Experimental Community Ecology: The Desert Granivore System, In: *Community Ecology.* J. Diamond and T. J. Case [eds.]. Harper & Row Publishers.

Citations within the text:

If you use information from primary literature, you need to credit the source. A proper citation will follow your first usage. Scientists seldom use direct quotes, so summarize the information in your own words. These "in test" citations are abbreviated, iusing only names and dates from the full citations in the references Section, and their format varies with the number of authors.

One author

Density-dependent regulation of prey population growth stabilizes the dynamic oscillations of predators and prey (Luckinbill 1979).

Two authors:
Simulations models suggest that the high-dose/refuge strategy can significantly delay the evolution of Bt resistance (Alstad and Andow 1995).

Three or more authors:
Human population growth reflects positive feedbacks, or "coalition effects" (von Foerster et al. 1960)

Peer Review Guidelines for the *Lemna* Report

Purpose: Peer-reviewing helps both the reviewer and the person being reviewed. The writer gets feedback on issues of content and clarity to help them make a higher-quality paper. The reviewer gets a chance to think about scientific writing from the point of view of the reader, which could help them clarify their own writing.

Peer Review Assignment:
Use the grading rubric for the assignment as a guide for what elements each paper should contain, and evaluate the paper in the following manner:

Skim the paper to get an overall impression of the main points. What is the author's hypothesis? Identify it by underlining it in the paper. In the margin, state whether you think the hypothesis is clearly stated, and any suggestions for clarification.

1. Read the paper more carefully and identify points of confusion by underlining with a wavy line.
2. Since this is a "canned" project, specification of the experimental design and rationale is quite simple; the underlying ideas, however, are tricky. Pay close attention to the author's portrayal of the issues addressed on pages 9 and 10. Are they well understood and clearly explained?
3. On a separate sheet of paper (typed), write your assessment of the IDEAS presented in the paper. Do you agree or disagree with the author's interpretation of the results? Do the papers cited, the figures and the tables support the author's conclusions? Does the author have sufficient data and background information? Are the references used appropriately?
4. In a second paragraph, describe two things that are particularly strong about this paper.
5. In a third paragraph, make three or four recommendations for specific things that the writer can do to improve the quality of the paper. Your recommendations may concern the organization of the paper, the content, the format, the readability or a recommendation about a repeated grammatical problem. For example, if the figures are of low quality, simply telling the writer to "make better figures" would not be a very helpful. Rather, you might suggest that they include their r-square value, label their axes or make a more detailed caption. Focus on the few things that could really improve the overall quality of the paper.

Adapted by Leslie Brandt and Angie Hodgson, from Bean, J.C. 2001. "Engaging Ideas: The professor's guide to integrating writing, critical thinking and active learning in the classroom." Jossey-Bass Higher and Adult Education Series, John Wiley, San Francisco, CA.

Detailed Grading Rubric for the Final *Lemna* Report

Biol 3408W, Spring 2008 *Lemna* Paper Point Breakdown

Each student will write his or her *own* report in the style of the journal **Ecology** (for examples, see Ritchie and Tilman 1995, additional articles at www.jstor.org). Suggested page limits (5 to 8 pages including text, tables and graphs, double spaced, Times New Roman font size 12, 1-inch margins all around) are just to let you know what we're expecting. More importantly, you must tell a complete "story" as described below. The total points as shown here add to 50.

Title (1 point) - Concisely describes research and its major finding(s)

Abstract (4 points) - Summarizes report in one paragraph.
Write this section *last*. Boil down your intro, methods, results, and discussion sections to a sentence or two each and you'll have a great abstract.

(1 pt)-succinctly describes why research was conducted in the first place

(1 pt)-succinctly describes methods (relative to all the possible methods in ecology - did you do a field experiment? Did you reanalyze others' data? In this large context, what did you do?)

(1 pt)-succinctly summarizes results

(1 pt)-succinctly describes societal or ecological implications of results

Introduction (10 points) - Provides background information on question. Describes hypothesis.
Write this section as if you haven't yet performed the experiment! Make NO mention of your results! This section sets the stage and lets the reader know why your question is interesting and worthy of pursuit. Part of that answer involves the scientific community and work that others have done on similar questions. At the end of this section, explicitly state your question and hypothesis.

(2 pts)-explains why this line of research is scientifically interesting

(4 pts)-reviews previous researchers' work (a good review of a few papers in greater depth or a good review of many papers in less depth earns 4 points. Between the introduction and the discussion at least 3 papers should be cited. If you select this lower bound, be sure to discuss the papers in some detail.)

(1 pt)-explains how student's work amplifies, tests, or complements previous researchers' work

(2 pts)-explains theory behind hypothesis

(1 pt)-clearly states question and hypothesis

Materials and Methods (5 points) - Explains what you did, why you did it that way, and what materials you used.
Write this section with enough information so that a peer could replicate your experiment (but not in a "first do this, then do that" kind of way). Organize it logically (which will seldom end up being temporally!). Do not include extraneous information (it can be assumed that you recorded information properly and kept track of your experiment properly - you don't need to write this!).

(2 pts)-explains and justifies experimental setup

(2 pt)-explains and justifies what data *are* collected

(1 pt)-explains and justifies statistics used in analysis (nothing beyond what we've learned in class is necessary) and any calculated values

Results (text) (7 points) - Presents all the data used to test the hypothesis. Summarizes important trends shown in tables/graphs (described below), and describes the results of statistical tests.
Write this section without analysis. Just say what happened. Do not "interpret" the results or put any "spin" on them.

(Imagine a researcher reads your paper years from now and has a radical new theory – one you've never dreamed of. She wants to see how your data conform to her new theory. She will use your results section to do this, and she does not want to be burdened by your interpretation ("spin") of the results. So just say and depict what happened!!!)

(3 pts)-properly describes in words the important trends depicted in figures and tables (failing to describe any important trend loses points)

(1 pt)-properly describes in words any important results that are not depicted in graphs (failing to describe any important results loses points)

(2 pts)-properly describes the *statistical* significance of the results

(1 pt)-*does not* discuss the implications of the results (*save that for the discussion!*)

Results (tables/graphs) (8 points) - Summarizes data visually (graphs) or "tabularly" (tables) along with appropriate statistical tests.

Graphs and tables should be as clear as possible with a meaningful title, labeled axes, and a legend that describes the graph or table and any statistics that are depicted in it. Each graph should have an associated figure number; the first graph referred to in the text should be figure 1, the second graph referred to in the text should be figure 2, etc. Each table should have an associated table number in a similar fashion. Don't imbed graphs or tables in your text; rather, include them at the end of your report. Note: it is permissible to print out a graph and then hand write on the graph: titles, axes labels, statistics, figure number, etc. Don't waste time wrestling with the software to print all those things out!

(1 pt)-all graphs/tables have a meaningful title

(1 pt)-all graphs/tables have well-labeled axes/columns&rows

(1 pt)-appropriate statistical information is included on graphs (p-value of test, R2 for regression)

(4 pts)-graphs/tables summarize the principle results meaningfully

(1 pt)- "legends" (or "captions") explain what each graph/table is depicting in complete sentences such that someone who just skimmed the main body text can understand the figure (might be slightly redundant with text, that's ok). Look at any published paper and you will see legends accompanying every figure – do this.

Discussion (10 points) - Interprets the results presented in the Results section in the context laid out in the Introduction section. Compares actual results with expected results. Discusses possible implication of findings in a scientific or societal context.

(1 pt)-discusses whether hypothesis was supported or should be rejected

(5 pts)-interprets the results, and discusses what the results mean. If methodology is discussed, it is done so constructively. (What we don't want is a discussion that reads "everything in our experiment was screwed up and so we can't say anything, blah blah blah". Something constructive can always be said.)

(2 pts)-discusses results in the context of other scientific work (as mentioned in the introduction) - do results support or contradict previous researchers' work? Why?

(2 pts)-speculates about greater ecological or societal implications of work and reasonably substantiates those speculations

Literature Cited (2 points) - List cited articles in the format used by the journal **Ecology** (use articles from **Ecology** as a template).

Citations in the body of the paper list the author and year (Lane 1998). Papers authored by two people use both of their names (Lane and BassiriRad 1997). Papers authored by three or more people only use the first person's name with "et al." (Dybzinski et al. 2003). Multiple simultaneous citings are separated by commas without using the word "and" (Lane 1998, Lane and BassiriRad 1997, Dybzinski et al. 2003). Notice that the period for the sentence goes after the closing ")".

Dybzinski (2003) claims that citations that include the researchers' name(s) in the sentence should have the year immediately following the name. Tilman et al. (2003) concur.

(1 pt)-citations are used properly in the body of the text

(1 pt)-the list of the papers cited follows the format of the journal Ecology.

Overall impression (3 points) - Lack of spelling errors, good grammar, good paragraph construction, and good overall organization.

(3 pts)-The paper reads "transparently". Poor grammar and organization do not obscure content. *Note that poor writing will likely lower your points in the sections above simply because your meaning will be unclear. Write well! Have someone proof read for you. "Tighten up" sentences that don't sit right.*

Modified by DNA 2007, L. Phillips (2006); Adapted by Ray Dybzinski (2005) from the work of Diana Lane.