The haiku of writing a paper

Joseph Craine
Updated 1/3/2009

General approach

Try to keep it simple

Broad brushstroke

My broad criteria for evaluating a paper are simple. First, a good paper should quantify an interesting ecological pattern. Second, the paper should test potential mechanisms for the pattern. These mechanisms should sit one level below the pattern of interest and provide insight into the processes that generate the pattern of interest. Third, testing mechanisms requires falsifiable hypotheses, preferably competing hypotheses, grounded in theory. Hypotheses should be clearly stated and be supported by introductory material.

Relationships among papers (From Terry)

1. Make a list of what stories (= papers) to tell and figure out what data goes with each story.

2. As you work on a particular story, make adjustments, pulling in some new data sets and deleting other SO THAT THE STORY CAN BE TOLD AS CLEARLY AS POSSIBLE.

3. As each paper gets formulated, there will be shuffling of data sets. It is probably worth keeping a list of orphan data sets and looking at it occasionally to think about how (or with whom) they might best be used

What data to include (From Terry)

I think your first criterion should be to use the data that tell a particular story most clearly. You don't want to leave out data because they contradict what you "want to tell". However, if a data set doesn't give a clear message because of other factors that are extraneous to your story (or which you didn't measure), that seems like a valid reason to leave them out. Judging from what you say, I'd be inclined to leave them out.

Abstracts

Start the abstract with most important results.

Do not end it with a statement that “results will be discussed” or “results have important implications…” Be direct on everything.

Introduction

Introductions and discussions are often similar.

Funnel from big idea to specific point

Standard three plus one paragraph model:

1) **Big question**. This is the general broad reference to your work. For example, it might be that atmospheric CO_2 concentrations are rising, or nitrogen is an important driver of ecosystem dynamics.
2) **Proximal question.** Within the broader framework, the proximal question should be stated that you are directly addressing. For example, although atmospheric CO_2 concentrations are rising, the controls over soil C storage are poorly known. Or although nitrogen controls ecosystem dynamic, there are important questions regarding the role of denitrification in controlling N availability. Note that a proximal question can often be framed as the big question—it's all a matter of perspective and how you want to tell the story.
3) **Scope of research with hypotheses.** In order to better understand the role of soil C storage in responses of ecosystem to elevated CO_2, we tested whether elevated CO_2 increased the C stored in the soil within soil aggregates.
4) **Competing hypotheses.** The best introductions and research designs test between competing hypotheses. Often when there is a single hypothesis that is rejected, the authors can derive alternative explanations that don't require the theory to be rejected. Therefore, might as well start with competing hypotheses since there are always competing hypotheses. Framing the hypothesis in the null form is not necessary when using competing hypotheses. For example, in testing the role of N in decomposition, an experiment could test whether stoichiometry predicted responses of decomposition to greater N availability. Or we can test between stoichiometry or N mining in predicting the responses of decomposition to greater N availability. No on experiment is generally able to reject a theory, so you can test between two theories and whether data supports one theory or another.

Methods

Past tense for methods

Include the basics of what would be necessary to repeat the project. Where things were done. How they were done. When they were done.

For equipment, cite the manufacturer and its location.

For statistics, basic statistics are not necessary. E.g. you don't have to state you did an t-test. Only when it is not completely obvious what test was done does one have to include a description of the statistical test. Include the statistical package and its version.

Results

Make the results more than just a list of findings. Include the biology.

[It is much more interesting to know that high root biomass draws down root nitrate than to tell them that a statistical relationship exists]

No references in the results section

All statements made in the results section must be associated with a statistical test.

Discussion does not belong in the results section although with judicious use of topic sentences, one can lead the reader to a discussion point occasionally.

[How to tell if a relationships is good, or "r^2 vs. P". Is a good relationship defined by the amount of variation that is explained or by the certainty of the relationship. Scenario with low r^2, but not good relationship. Scenario with low r^2, but good relationship with low P.]

How many different ways to show data, especially if different results arise from different approaches to analysis. On the one hand, concise papers will use the best methodology to test ideas. At some point, readers will have to trust the authors. On the other hand, it could be argued that there are multiple ways to analyze data and readers should either be shown that multiple approaches lead to the same conclusions, or that there are some equivocations that arise from different assumptions. There can be no one universal approach to the dichotomy. Yet, if there is one best way to analyze the data and nothing is gained by showing multiple approaches that lead to the same conclusions, then concision should override certainty. If there are differences, then authors should decide between competing approaches as to the best analysis.

Discussion

From Terry: *There are at least two possible strategies: (1) Make the structure of the discussion parallel that of the results—i.e., start out with seasonal change (probably the easiest approach for readers to follow) or (2) start out with the most important result first so that the reader will be sure to get engaged in the discussion. I'd be inclined to use the first approach, but to start out with the most important point about seasonality; then go with the most important points about size effects on root parameters.*

Four components of the discussion

1) **Proximal interpretation, comparison of results to hypotheses, and cross-relationships among findings.** Sometimes, when writing the results section, one is unable to fully interpret the results. Some things might not be clear from a simple reporting the of the data. A and B might be positively correlated, but the ecological significance of that might not be clear. It is best to write the results section so this doesn't happen. If it does, interpret the results more fully here. In addition, in the results section, it is often hard to link the various results together. Then, make sure explicit hypotheses are discussed in relationship to the results.
2) **Relationship to findings in other papers.** After the reader is clear about the proximal interpretation of the results and how the results relate to one another, it is important to compare results with the findings of others.

3) **Relationship to broader theory.** Once the results are set in context with regards to the findings of others, it is important to now reexamine the concepts and theories on which the paper was based.
4) **What needs to be done next.** This section should be short. Very short. Discussing what should be done next is the most abused part of discussions. That stated, sometimes research results, especially if conflicting, narrow in on a key question or measurement that needs to be done. If this is the case, it should be stated.

You should not introduce any new data in the discussion. That is for the results section.

Citing

Figures

Never have a figure, where a table will work. Because tables have the actual numbers, they are superior to figures. All bar graphs should be shot and tables put in their stead.

Tables

Punctuation

Use an n-rule (n-dash) to separate entities (photosynthesis–N relationship), but a hyphen for others (N-supply gradient)

"e.g." should be followed by a comma

PCAs is the plural (no apostrophe any more when using capitalised acronyms)

Grammar

Data were, not was.

among vs. between

Strata vs. stratum

Taxa vs. taxon

Function not functionality

With respect to, not with regards to

Consequently, not as such

Not something deep, but at depth

Replace space-space with space, except colon space space.

Search for space-comma or space-period and remove.

Do not use measurement when you mean property/characteristic/metric

replace basic with general

Than vs. then

Think about five vs. 5: if exact number that you would measure exactly, use numeral…except if first word sentence

Among not amongst

Use [phrase], and [phrase] rather than [phrase] and [phrase]

Commas, semicolons

Hyphenated adjectives. For example, “low-N plants”, not “low N plants”. Also, “least-squares regression model”

Avoid the naked this

Turnover is a noun, turn over is the verb

Nouns modifying other nouns: plant traits vs. traits of plants. Applies to “by” and “for”.

…that were observed vs. observed

Alternative hypotheses contrast one another. Alternate hypotheses would follow in sequence like left right left right. Use alternative or alternatively.

Do not split an infinitive verb with an adverb, but it is OK to split an auxiliary verb from a verb. “To lazily read” should be “to read lazily”. “was consistently limited” is OK as is “consistently was limited”.

Style

Be direct on what you did, never attempt…

How to report results

Models permitted, not included the ability

State what happened directly. Do not be vague. Do not state that “treatments varied” when you can state that treatment A was greater than treatment B.

Other

Due to sample variability, not most likely due to noise

Don’t use “as mentioned earlier” if you can help it.

“can”, not “is able to”

were, not were determined to be

Don’t talk about the models, just the results.

Results of the model are past tense.

Check on correspondence between intro, results, and discussion—both for order and 1:1

Make the experimental process transparent. Not we found that x increased, but x increased.

It’s not a pot, it’s a container.

It’s not a microcosm, it’s a mesocosm.

Things aren’t “characterized by having”, they “are”

Use provide, not serve the purposes of providing

Do not say shown, say what it is

Do not say that it is known to do something, say what it does

replace amount of x with x. e.g. amount of fine root biomass becomes fine root biomass

No then we did this

Rarely, not almost never

“Which [verb]” replace with gerund. E.g. “which reduces the power” should be replaced with “reducing the power”…

Avoid use of “it” and “there” as subjects of a phrase or sentence. These words are both indefinite pronouns, and the goal of science is to be more definite! Often phrases can be rewritten with little effort, and thus result in sentences that are clearer and shorter. (New Phytologists reviewer)

Avoid compound sentences with compound subjects, verbs, or objects. These types of sentences are very “awkward” and difficult for readers to understand.

Christine Bezar: how to refer to negative numbers that are greater or lesser than another negative number. I think he agrees that more or less negative is a no-no, but that trying to describe the situation to a reader can end up being very wordy and off-putting. I know of no conventions in this so don't know how to support your case to the journal (Maths is not my strong point so I rely on the scientist peer reviewers for this). In maths are you allowed to say 'Both A and B were negative but A was closer to zero than B' ? In English I would have thought you linked lower with higher, lesser with greater and not mixed them, but in Maths maybe mixing is allowable?

Technical Conventions

Table and Fig #’s in order of appearance

Before you finish, check correspondence between all data reported and tables and figures.

Also check to make sure that all results are contained in tables and accurate.

Use nitrogen the first time, place in parentheses, and then abbreviate with N. Same for carbon.

Plagiarism

Do not copy text from another person’s work.

Self-plagiarism to a degree is allowed in the Methods section. It's OK to use the same text for the same methods in different papers, but don't overdo it. It's not OK to use the same text in any other section.

Unpublished vs. not shown

"Craine unpublished" suggests that you intend to publish the data eventually. "data not shown" implies that these measurements are part of this study but that you aren't showing them to save space

Use of subheadings

Sentences a thought

Paragraph

subheadings

Conventions on reporting data

Italicize r^2 = and P =

Report r^2 to two decimal places.

P values should be exact, except for $P < 0.01$, <0.001.

Joe-specific

Coarse belowground biomass, not coarse roots if the fraction contains rhizomes.