Basic outline and suggestions for writing an empirical academic research paper
By Tom Carsey (carsey@unc.edu)
Last Updated: 2-18-2016

(NOTE: I think these points generally apply to both qualitative and quantitative empirical papers, but I assume a quantitative paper below to keep this relatively simple)

It is presumptuous for anyone to offer advice about anything. I am confident if anyone strictly applied the advice that follows to my own publications, they could keep themselves busy pointing out all of the times I failed to practice what I'm preaching. Still, from my own experiences writing, advising dissertations, theses, etc., grading endless seminar papers, and serving as the Editor for *State Politics and Policy Quarterly* for four years, I do have some thoughts on how to assemble research papers. I start with a detailed outline for a standard empirical paper, and then offer some additional thoughts.

1) Introduction (2-4 paragraphs) – must include the following
   a) What is the problem, puzzle, tension, etc. motivating the paper?
   b) Why is providing an answer to that problem important?
   c) What is the answer to the question, puzzle, etc. you posed above that you are going to test in this paper? Why is this the answer? (Answering "why" is the core of your theory).
   d) Sketch of how you are going to test your answer: "in this paper, I will test this theory in the context of U.S. state legislative elections" (or whatever).
      i) NOTE: parts 1a, 1b, and 1c should be written without specific reference to the context in which you will be conducting your test. For example, if you are testing your theory in U.S. state legislative elections, your paper should be framed in more general terms – legislative politics, representation, elections, etc. You have a general question that you are testing in a specific context.
   e) Optional components of the introduction:
      i) An anecdote, historical event, etc. that motivates the problem. This can be a nice hook for readers, but resist the temptation to run on.
      ii) A final paragraph that outlines the rest of the paper (e.g. first theory, then data/methods, then results, then conclusions). I personally find these paragraphs to be a waste of time and space, so I recommend not having one.
2) Theory/Literature/Expectations
   a) This section is really designed to unpack 1a, 1b, 1c, and 1d. Let me offer more detail.
      i) Unpacking 1a is critical, but probably should not be the longest subsection of this section of the paper. Cut to the core.
      ii) Unpacking 1b should not take too much time either. This is something to which you will return in the conclusions.
      iii) Unpacking 1c – your articulation of the theory that provides an answer to your problem – is extremely important. This is your argument for how the reader should understand why the political process you are studying works the way it does. This requires being clear about assumptions, how those assumptions fit

together to produce an argument, and in general terms what the expectations/observable predictions are from that argument.

       iv) Unpacking 1d requires some detail on the context you are using to evaluate your theory. This portion probably merits its own subheading.

          (1) You need to discuss the pros and cons of testing your theory in this context.

          (2) You need to translate theoretical concepts into variables (at the conceptual level – not specific measures yet).

          (3) You also need to translate your theoretical expectations into specific hypotheses.

b) Some general comments on this section:

    i) Others disagree, but I do not think there should ever be a section heading called "Literature Review." Your job is not to review the literature. It is to make an argument. That argument is informed by the literature (and citations to those points are necessary). You also need to place your argument into the larger debates/findings in the literature. Thus, the literature matters in a critical way, but your job is not to just summarize what others have done. Those summaries make for dull prelim answers and even duller research papers. I encourage building an argument like building a puzzle. The literature offers some of the pieces, but your job is to put them together into the picture (argument) you want to make, filling in the holes with your own theoretical contributions.

       (1) Do not even refer to the literature in your writing. By this I mean a sentence like, "Race strongly predicts voting behavior (Smith XXXX; Jones XXXX)" is much better than a sentence like, "The literature finds that race is a strong predictor of voting behavior (Smith XXXX; Jones XXXX)." Just make relevant declarative statements and provide citations.

    ii) Having an orderly flow to your argument is critical. If you lose people in this section of the paper, your analysis, findings, and conclusions will not make sense to anyone.

3) Data/Model/Methods

a) You already told them about the context in which you are testing your theory in 2.a.iv. Here you need to talk about the specific data you have.

b) You need to operationalize the variables.

c) You need to specify the statistical model (write the equation(s) or at least some form of it/them).

d) You need to say how you are going to estimate the parameters of the model.

e) You need to offer specific predictions about findings from the analysis that would constitute evidence supporting your hypotheses. You have now completed the translation of your general theory into specific predictions about what you will observe in the data if you are right. Getting this translation right is absolutely key.

    i) NOTE: you also need predictions about any "control" variables you include as well. You do not need to devote a lot of space to them, but if you don't have a prediction, you have not offered a rationale for why that control is needed in the first place.

f) You should also acknowledge any problems with your data, measures, or methods. Serious issues should get discussed in the body of the text. Minor issues are better dealt with using footnotes.

4) Analysis/Findings

   a) Document EVERYTHING you do in a STATA Do file, an R script file, or in whatever software you use. Write that file from opening the raw data through all coding/recoding, to all measurement choices, to all modeling choices. Fill those files with comments that explain every step along the way. If you cannot hand any person the raw data and a file that works from that raw data to produce every table, figure, and test you conducted and presented in either the paper, an appendix, or in a footnote, you have not properly documented your research.

   b) Every paper needs a table of descriptive statistics for the variables that are used in the analysis. For some variables, it might be the mean, standard deviation, minimum, and maximum values that are needed. For categorical variables, frequencies are more helpful. It makes no sense to consider relationships between variables in a more complicated statistical model without first knowing descriptive information about each variable individually.

      i) A common mistake here is to present a table of descriptive statistics for variables that does not take into account cases that are later dropped from the analysis. If you report a sample size in Table 1 that is larger than the sample size of the regression (or whatever) models you estimate later, then Table 1 is not providing the proper descriptive statistics of the proper portion of the data.

      ii) Of course, dealing with missing data is itself a significant topic that merits explicit attention to its presence and your response.

   c) Presentation of the main model. There are many issues to consider here:

      i) Some like to present a series of models working up from a baseline model to a more complete model. This can be informative sometimes, but it also raises the question of whether it makes any sense to interpret the baseline model if a later model is being offered as the one that is correctly specified. My default is to only present the correctly specified model unless a compelling case can be made to do otherwise.

      ii) At some point, you need to do a complete set of diagnostics regarding your analysis. Are their problems with multicollinearity, hetroscedasticity, serial correlation, endogeneity, unit roots, unit effects, influential outliers (or clusters), measurement, model specification, selection bias, clustering in the data, missing data, etc. You want to be the one to find problems with your paper before some reviewer or later author does. You want to conduct all necessary diagnostics and report them.

         (1) You should do so in a systematic way, record all of the things you did in a Do file in STATA or script file in R (or whatever), and provide all of that as an appendix to the paper. Generic footnotes saying "I checked and this was not a problem" without any documentation or specific reporting won't get passed most reviewers/readers anymore.

      iii) Once you are satisfied with your diagnostics, you should interpret the results of your model.

         (1) Start with the key findings relevant for evaluating your theory.

(2) Sign and significance are not enough. Provide a substantive interpretation of the magnitude of effects you uncover. Present meaningful quantities of interest.

(3) Provide a clear representation of your uncertainty. Just reporting point estimates and stars alone doesn't cut it anymore. Give the reader a sense of the distribution of plausible estimates (e.g. confidence intervals – or credible intervals if you are Bayesian), or just show the distribution of the estimates.

(4) Use tables where that is best, use figures where that is best. Some say figures are always better than tables, but I disagree. Figures are often better than tables, but not always. Think carefully about how to do this, and ask your friends to weigh in to see if they understand your tables and figures.

    (a) All tables and figures must be able to stand on their own – use clear and complete titles, notes, legends, etc. Do not use variable names that do not mean anything to readers (Write out "Logged per capita income" rather than "lnpcinc".

    (b) Format tables and figures like you see the best ones presented in journals like APSR. Nobody wants raw output or a cut and pasted STATA figure.

(5) Make sure in the body of the paper that you direct readers carefully to where they can see findings in tables or figures as you describe your results in the text.

(6) Be explicit about whether the findings you have are or are not consistent with your theoretical expectations.

(7) Provide some discussion of findings related to control variables, but be brief.

5) Conclusions/Discussion

    a) People disagree about whether there should be two separate sections or just one. I tend to prefer just one, which I generally label "Conclusion."

    b) Do NOT spend a lot of time restating your empirical findings. No more than one paragraph that summarizes the collection of findings is needed, and sometimes not even that.

    c) Instead, link your findings back to your theoretical expectations at a conceptual level. Is there strong/weak/mixed/no support for your theory?

    d) Relatedly, what are the implications of your findings for how readers should think about the problem or process under study? Return back to the introduction – did you solve the puzzle? What does this mean for how we might think about other processes? Are their normative implications stemming from your findings? Are there implications for other areas of study (e.g. some studies of voter behavior have implications for candidate strategy)?

    e) What problems or issues arose in the study that you think are important for readers to consider?

    f) Where should research in this area go next?

        i) There is disagreement on whether this "future research" bit belongs in most papers. I'm indifferent on this issue, but I do think you should avoid lengthy

descriptions of future research. Such discussions lead readers/reviewers to say, "If these are such good ideas, why didn't the author do them already?"

Next, I present a list of several general comments about academic writing. They are not in any particular order – just some thoughts as you work on the craft of writing good scholarship.

1) Your primary goal is to communicate your thoughts and choices to the reader clearly, precisely, efficiently, and effectively. Specialized jargon, technical terminology, and dense writing should be avoided. You do not want to hope readers think, "I don't understand this, so I guess it is smart." You want them to think, "I understanding this completely and it sure is smart." Most papers would be better if they used shorter words, shorter sentences, simpler sentences, and fewer overall words.

2) All drafts should include page numbers.

3) First drafts are never acceptable final drafts
   a. The good news is that this realization takes all the pressure off of writing the first draft. You know it won't be the final draft anyway, so there are minimal penalties for errors.

4) Related to the previous point, editing involves more than making sure spellcheck is turned on. Serious editing engages every word, phrase, sentence, and paragraph to see if it serves the primary goals outlined in point (1) of this list. The number of revised versions might easily reach double digits.

5) One last point on editing – don't expect reviewers, editors, advisers, etc. to proofread for you. Clean up grammar, spelling, punctuation, etc. as best as you can yourself, get help from peers, or hire help. If you ask an adviser to read a draft, let them know if you think it is "clean" or not in advance.

6) Never send out a paper to a conference or for publication that is not clean, proofread, and ready for public display. Make publication quality tables, figures, formatting, etc. Your work should reflect your seriousness and your professionalism. Reviewers hate sloppy submissions and they penalize sharply for it.

7) You should prepare a detailed outline of your paper at some point in the process. Some do so before they write a first draft. Others wait until after a draft or two. Either way, creating a detailed outline will impose discipline and structure on your final draft. It will help you find gaps, redundancies, and problems with the order of elements in the paper. It will also help you avoid tangents and digressions.

8) Give serious attention to your introduction and your abstract, but don't worry too much about it for the first draft. It is easier to introduce something after you know what it is you are trying to introduce. Furthermore, the abstract is really meant to be the thumbnail of a completed paper, not the introduction to a paper, so write it at the end. Still, the abstract and introduction are the first things eventual readers will read, so they need to be clear, tight, and

persuasive. In my experience, the opening paragraph of most first drafts is best deleted at some point as it often presents a vague and general "Democracy is important" sort of claim.

9) I am <u>not</u> a fan of:
   a. Rhetorical questions – they waste space because you need to stop and answer them.
   b. Argument by metaphor
   c. Argument by analogy
      i. Both metaphors and analogies are weak forms of argument. Say what something is rather than what it is like or similar to.
   d. Double negatives: especially sentences and phrases that say things like "not only does . . ." Telling the reader what something is provides infinitely more precision than does telling the reader what something isn't. (What time is it? Well it's not 11:30?)
   e. Long block quotes from the literature (though I can see exceptions)
   f. Long wordy sentences with lots of (often improper) use of punctuation.
   g. Words, phrases, sentences, or paragraphs that do not convey something that is essential
   h. Long papers
   i. Unnecessarily negative tone
   j. Starting sentences with conjunctions like "And," "But," or "Or."
      i. Bad grammar generally
   k. Efforts to be cute, flowery, or clever because:
      i. Such efforts almost always fail
      ii. They are almost always imprecise
      iii. They waste space
   l. Using the phrase "in theory" in a paper. It should be obvious when you are writing about a theory. Also, in non-professional language, people often use the phase "in theory" to suggest that while something is supposed to happen one way, it often happens in a different way. That is not what we mean by theory in our professional writing, so using this phrase just confuses people.
   m. Using the word "significant" to describe anything other than whether a result is statistically significant. Reserve the word "significant" for statistics and pick something else when you want to say something is important, consequential, meaningful, etc.
   n. Writing sentences that reference the literature – "the literature finds …" Just write declarative sentences and add citations as appropriate.
   o. Putting quotes around words as if you were speaking and using so-called air quotes. Putting quotes around a word (unless you are clearly just using them to denote a label) generally implies to the reader that what you mean is something somewhat different from the actual word you are using. In fact, it often communicates to the reader that you mean something intentionally different from the actual word, though what exactly you mean is left unspecified. Again, this is just another form of imprecision in writing.

10) My (nearly) last tip is that the best way to write well is to write early and often. We can only organize so many thoughts in our heads. We can talk with friends and colleagues about our work, but it is easy to miss or gloss over gaps and inconsistencies in our thinking when we are only thinking or talking. Writing imposes discipline and structure on thinking. Writing anything at least gets you past that blank screen and started toward producing something. You should have high standards for your final drafts, but low standards for first drafts – just get that first draft started. You know you will be changing it so there is no pressure to make the first draft perfect or really even that good.

11) My actual last tip is to share your writing with others. Trusted colleagues, advisers and the like are a good first audience. Get use to constructive criticism – in fact, seek it out. The friend who makes you feel good by saying everything you write is awesome is not really helping you improve. This is a business where we put ourselves out there for public critique all the time (submitting papers to journals, presenting at conferences, teaching, etc.). You have to find that balance between confidence and being open to criticism. You can't be too sensitive to criticism, but you also can't be close-minded or defensive about it or you won't learn anything from it either. None of us are so smart that we can't learn something from others nor so ignorant that we don't have something to contribute ourselves.