

Comment

Are we training pit bulls to review our manuscripts?

Virginia Walbot

Address: Department of Biology, Stanford University, Stanford, CA 94305-5020, USA. Email: walbot@stanford.edu

Published: 9 March 2009

Journal of Biology 2009, **8**:24 (doi:10.1186/jbiol125)

The electronic version of this article is the complete one and can be found online at <http://jbiol.com/content/8/3/24>

© 2009 BioMed Central Ltd

Abstract

Good early training of graduate students and postdocs is needed to prevent them turning into future generations of manuscript-savaging reviewers. How can we intercalate typical papers into our training?

Who hasn't reacted with shock to a devastatingly negative review of a manuscript representing years of work by graduate students and postdoctoral fellows on a difficult, unsolved question? Detailed in its critique, it relentlessly measures the work against a 'gold standard of excellence' using the latest and best techniques, before dismissing the years of labor and stating that the manuscript can only be reconsidered with substantially more data providing definitive proof of each claim. The other two reviews may be favorable - even recommending publication with few revisions - but how can an editor ignore that complete and negative review? Your manuscript is declined, with encouragement to resubmit when new data are added.

I confess. I'm partly responsible for training the pit-bull reviewer, and I bet you are too. Graduate students read, discuss and dissect classic papers as a key part of their training. At Stanford, these discussion sections are led by faculty. The 'best practice' papers chosen for close reading provide training in how to frame a question, how to mine the literature for relevant biological materials to conduct new experiments, and how to construct studies with appropriate controls and analyses to extract conclusions. Faculty ask students to summarize the article's claims, gleaned from the abstract and discussion, and then to judge the quality of the evidence for each claim by a careful reanalysis of the data. Some of these papers have been the turning point in a field or the first in a field - papers completely worthy of this exercise.

We also teach using papers, published in prominent journals, that contain fatal flaws, not fraud, just faulty assumptions about the properties of organisms or reagents, lack of appropriate controls, or a failure to consider alternative interpretations or to mine the literature completely. A favorite in plant biology is a paper claiming massive and dynamic movement of sequences from the mitochondrial into the nuclear genome, followed by amplification of these mitochondrial sequences - perhaps in the manner of transposons. The paper opens with the statement that plants contain three genetic compartments: nucleus, mitochondrion, and plastid. Too bad the authors, the reviewers, and the editors did not take this instructive sentence to heart. All of the data are DNA blot hybridization assays depicting wide fluctuations in hybridization of a particular probe to the nuclear fraction, with mitochondrial hybridization constant. Students reading the paper identified a key 'missing' control, namely inclusion of purified plastid DNA. In fact, further work showed that there was a historic transfer of a tRNA gene from the plastid to the mitochondrial genome; hence the study had been tracking relative plastid DNA content (a type of contamination) in nuclear DNA samples.

There's nothing wrong with using either classic or fatally flawed papers in our teaching, provided we also instruct our students about what constitutes a more typical publication. Few of us will ever write a classic paper - the simply outstanding paper that might garner the authors a Nobel Prize or provide a completely surprising new insight or a

significant new technique. The papers that represent great leaps forward are few in number. And we all work to avoid submitting manuscripts with fatal flaws - the internal review of lab group meetings and colleagues is designed to avoid horrible mistakes.

The majority of our collective publications, and hence scientific progress, comes from incremental insights in which the context is provided by the ongoing struggle to resolve a number of outstanding questions in a field. A series of papers, often from different labs over a span of several years, will add up to the solution to one or several questions. Each publication was timely when published, but may be wrong in some of the details of interpretation - the focus in the discussion may have dealt primarily with the most popular model, missing the chance to 'redesign' that model to better fit all of the data. None of these papers is a complete answer: the new insights will eventually be summarized in a short review article weaving the incremental threads of data into one story that becomes the new paradigm, at least for a while.

Taking a phrase from the current US political scene, these experimentally solid papers are "timely, targeted, and temporary". That is, they address unanswered issues that are on the minds of those in the field, they target specific issues amenable to experimental or theoretical resolution, and in some ways their impact is temporary, because subsequent papers using the emerging insights and new methodologies will supersede these solid papers. Yet these solid papers are the foundation for progress most of the time.

Students are trained to be pit bulls in finding even the tiniest faults in great papers. Nearly all the truly remarkable papers we teach contain a few 'typographical' errors such as reference to the incorrect panel of a figure or a small mistake in a large table or the wrong initials for an author in the reference list. These errors do not detract from the impact of the work, but instruct students to be vigilant in that even the deservedly famous can make mistakes. This insight may even inspire some students to use spell-checker and other automated tools to eliminate such errors. Similarly, the papers with fatal flaws, particularly those in which a critical control is simply missing, are highly instructive. These papers highlight the dangerous 'snow globe world' of belief in a particular theory - a world circumscribed to consider only those things within view - and even then only when obscured by snow. It's instructive to point out that the meaning of 'belief' is to accept as true in the absence of facts. The papers with fatal flaws help students appreciate that maintaining skepticism about current interpretations is essential for progress.

How then can we teach students to appreciate the bulk of our own contributions to the literature? Great manuscripts with minute flaws and bad papers with fatal flaws will represent a tiny minority of the manuscripts that our fledgling reviewer will actually encounter. The majority of manuscripts will be sound in conception and fair in data presentation, and contain some new information. How do we teach judgment of where in the pantheon of journal quality a particular study belongs? How do we teach what constitutes a timely 'publishable unit' - not complete proof of a major concept but a defined step in that direction? Here are a few suggestions - ideas that I hope will start a conversation about training reviewers and better scientists.

Read a short review and all of the constituent papers to understand how solid, but as yet incomplete, papers add up to a new paradigm

The class should discuss where these papers were published. Which made it into *Science* or *Nature*? Which were in the most visible biology journals and which in more specialized journals? Were any in obscure journals but cited by others in the field? Rosalyn Yalow, co-inventor of the radio-immune assay (RIA) technique and 1977 Nobel laureate in medicine, opened her seminars in the 1960s with the statement that the original manuscript describing RIA was rejected in all the best places, then in the not so good places, and finally found in a home in the *Journal of Clinical Investigation* (which at the time was well down the pecking order). Early citations were self-citations, but the quality of the journals she published in gradually improved and then the world discovered what you could measure, and papers in all the best journals used the procedure.

Points for discussion on this topic would be:

What were the claims and evidence in the papers cited in the review? What constituted a publishable unit in this field, at that time? Is there a substantial difference in quality between papers in the most prestigious journals, in specialty journals in the field, and in obscure journals? In retrospect, given the emphasis in the review article, are the key conclusions primarily from the papers in the best journals? That is, did reviewing at the time identify the papers that best established new points or clarified existing concepts?

What models or accepted ideas were being examined in greater detail in the suite of publications? Was the final answer the proof of this model or did a new paradigm emerge with the unfolding of the story and incremental data?

Did understanding await invention or implementation of a new technique?

Did information from another field - outside the general scope of the suite of publications - alter thinking substantially? Did new resources such as the publication of a genome or protein interactome from a high-throughput science project provide essential information for the field as whole? Would these new data types have been generated by the individual labs in the field, given their resources and expertise?

Listen to short presentations from several graduate students and postdoctoral fellows from one lab (equivalent to a faculty research seminar in depth and breath) and then discuss what's ready for publication

Should this 'story' be one publication? Or can the work be broken into distinct publications? Should it be broken up? In a perfect world, what additional information would be obtained before publication? Which claims (conclusions, hypotheses) in the story have strong support and which are new ideas, perhaps with little direct support? The purpose is to have students consider what constitutes a timely publishable unit of information in a particular field and how the ongoing contribution of new ideas and partial proofs stimulates work in the field.

Who will be an author? If there is just one publication, which dataset merits first authorship? The purpose is to discuss the realities of authorship, the need for both students and postdocs to have 'rights' to their own work, and the impact on careers of a single publication in which most of the participants are *et al.*

Ask the lab to provide a timeline of when particular projects were started and what tools or new information became available during the project and whether these were incorporated into the study. The purpose of this exercise is to teach realism when reviewing: were the questions posed and the methods used timely and updated appropriately within a reasonable span before submitting the manuscript?

As a class exercise, discuss how the project would be formulated today given the 'best techniques' and available information. Compare reality to a design that can take advantage of all new information and techniques available.

Compare the costs of the actual path to information and the best possible approach, both in terms of human effort and materials. Would the best effort require a genome project or other large-scale effort outside the scope of most labs?

Consider the possibilities of partnerships to conduct the best possible study versus individual lab efforts (even individual people efforts). Would the field be best served by

waiting for funding for the 'best' project? Would training be better served in individual or large group projects?

If those submitting manuscripts are honest - and most of us are our own best critics - about the timeliness and completeness (given constraints of time, effort, funding) and share the intent to make a solid contribution on an important question, then what we ask of reviewers is that they consider this context in writing the review. Sure, it's easy to trash a manuscript missing a paper published online this week or that fails to spend a million dollars to get a proteome of the cell types in question - but is this realistic? The trend to read manuscripts in PDF format on a screen also means that it's tempting to just start typing comments without first considering the manuscript as a whole - perhaps the issue so bothering you 'right now' is actually addressed in a subsequent section, perhaps even in the Materials and methods, now shuttled to the end of nearly every manuscript. With paper manuscripts, most reviewers read the entire thing - perhaps dragging it around town for days - and then sat and composed a review that had the perspective of a complete reading. Those old enough to remember paper manuscripts arriving in bulky packages in the mail may have learned better habits of scholarship imposed by the medium. Now it's up to all of us to teach 'best reviewing practices' to our students and postdocs and to use them ourselves.