EDITORIAL

Data dredging, salami-slicing, and other successful strategies to ensure rejection: twelve tips on how to *not* get your paper published

Geoff Norman

Published online: 29 January 2014

© Springer Science+Business Media Dordrecht 2014

Everyone talks about "star systems", like professional sports, movies, even airline piloting, (and now more recently the financial industry) where a few at the very top make scads of money, and many at the bottom just scrape by. It seems to me that we in the academy are also players in a variant of a star system, where many try very hard but relatively few succeed. The system is university research, and the currency is publications, not money.

Consider the fate of a research idea. Far too commonly, the idea arises in response to a practical question, where someone reflects on her educational roles and identifies a question that she apparently has no answer for. "Are expert tutors better than non-experts?" "What is the optimal size for a tutorial group?" "Do multiple choice and short answer questions give different information about a student?" "How can I accommodate students with different learning styles?" "Should I use periodic quizzes to reinforce learning?" Far too often this then leads to a study, where she uses her skills as a clinical researcher to design some research to address the question. And far too often it emerges that the study, while methodologically acceptable by clinical trial standards, is missing some critical elements for good educational research. Or, as commonly, it is a question that has already been answered.

Along the same lines, far too often, a faculty member with acknowledged educational skill but minimal interest in research will sit down with his department chair for annual review, and be advised to "Write up that course you're doing and get it published." A futile exercise; very few journals in education will publish full articles that are little more than descriptions of bright ideas.

Finally, there is the research requirement that residency programs and specialty boards demand of residents. Far too often residents will be required to conduct a research project, all by themselves, with minimal supervision, in their spare time. Often this is some kind of educational research, typically a survey of other residents, since it looks easier—no patients, no ethical issues. This is a recipe for mediocrity, and likely does more harm than good in turning residents on to the value of educational research.



McMaster University, Hamilton, ON, Canada

e-mail: norman@mcmaster.ca



2 G. Norman

Why do I say "far too often?" Does this reveal a certain arrogance on my part? I think not; rather, it reflects the nature of the field, where many of these research endeavors are likely to lead to frustration and failure. The reality is that only about 13 % of the papers submitted to Advances are accepted, and of those that are rejected, only about one in three is published elsewhere. I expect acceptance rates at other journals like Academic Medicine and Medical Education are, if anything, lower. This is not a conspiracy, nor is it based on a quota system. Rather, we accept 13 % of submissions simply because only 13 % of submissions are acceptable.

This represents an enormous waste of resources—fiscal, physical and intellectual. And this is only one hurdle. I have not considered the low success rate on grant applications, the number of funded studies that are not completed and then the number that are not submitted for publication. I cannot change the world, but I might have some insights about why success in publishing is so low. From my position as editor for nigh on 20 years, I see a number of things that authors could do at the design, execution, analysis and write up stage to increase acceptability. Here are some suggestions, framed in the negative as strategies that are bound to fail:

(1) Nice ideas finish last

There was a time when educational research had its origins in the grassroots, and we PhDs viewed our role as providers of methodological support for the clinicians who were the source of the ideas. This is no longer the case, as I have written elsewhere (Norman 2011). The field has matured and many academics, with solid disciplinary credentials (which I don't have) have entered the field. It is no longer acceptable to begin with a problem statement, and provide insufficient theory or evidence to justify the study.

(2) The world doesn't need another mousetrap

Whatever the down side of it, we rarely deliberately replicate studies. Far more commonly a study gets replicated simply because the author was unaware of previous work. It is now the case that many manuscripts are rejected because the literature review was incomplete or inadequate. A good literature review is a sine qua non. This should arise early in the development of the idea, as it then permits refinement (or abandonment) of the study in light of the evidence available.

It's strange, but courses on research design spend most of their time on how to get a good answer to the research question, but very little on how to get a good question. Perhaps Karl Popper said it best: "Good research amounts to refuting cautious hypotheses and affirming bold hypotheses". But you only know what's bold if you know what has gone before.

(3) We don't need to compare something with nothing

While the debate continues to rage in clinical research about the merit of a placebocontrolled drug trial versus a comparison to conventional therapy, the debate is long over in education. We'll accept without proof that some education is better than none. Cook, for example, has done a number of systematic reviews (2008) of educational interventions showing an effect size of one against nothing and zero against anything.

Note that we're not simply talking about an experiment where one group gets the cool new intervention and a second group gets nothing. This "something to nothing" design arises in many different ways. For example, when you look at a study where your class had a pretest, some instruction, then a posttest, and showed a large gain, you've compared something (after) to nothing (before). When you compare the class at the end of the year to



its performance at the start of the year, you've compared something to nothing. Any of these many one group—pretest–posttest designs are called "pre-experimental" by Campbell et al. (1963) and are viewed as pretty well useless (and were when the book was written in 1963).

(4) A + B is always greater than A

Another variant on research design that proves nothing is the A versus A + B design. It arises frequently in studies of gizmos like iPhone apps, or PDAs or simulations, where one group gets the standard instruction and the other group also gets to listen to heart sounds or view ECGs on their phones on the way to class (if they can hear the heart sounds above the roar of Guns 'n Roses). Just as we need not prove that something is bigger than nothing, we also do not need to prove that something + something else is greater than something alone.

(5) We don't really care if your students loved the course

Students don't know what they know. Studies of self-assesment (Eva and Regehr 2005) have consistently shown that self-assessed abilities are uncorrelated with actual performance measures. Since this is the case, it makes no sense to use self-proclaimed achievement or satisfaction ratings as an outcome in a curriculum. AHSE has a policy that we will not accept any study that uses self-assessment as an outcome measure for good reason. It tells you very little about the outcome.

(6) We are scientists, not inventors

Particularly in simulation studies, it is very common to show that a particular simulation leads to a learning gain (Issenberg et al. 1999). All too frequently, this is compared to nothing or to "conventional approaches" (which generally amounts to the same thing). This may be of value if your company is marketing the particular device. But these studies suffer two glaring defects. First, they are comparing something to nothing (See points three and four above), which is only interesting if something doesn't win. Even when you compare one to another (Norman et al. 2012), pricey simulations rarely win over other simulations, even when there may be a factor of 1,000 in price. Finally, these case studies are really just market research—what Cook (2005) calls "media-comparative studies". Without knowing a lot more about the constituents of the simulation, there is little of general interest unless you are dead set on buying this and only this simulation.

(7) The only people who care about education of pediatric gerontologists in Lower Volga are pediatric gerontologists in Lower Volga

Far too many papers begin with a statement about problems associated with the training of health professional X in school Y or country Z. Unless there is some reason to believe that these problems are much more widespread, then the article belongs in a national, not an international, journal (or no journal). As a corollary, it is almost inevitable that articles in this genre will inevitably tell the reader far more than they care or need to know about the particular bureaucratic and political environment in country Z, inevitably sprinkled with multiple three or four letter acronyms of various agencies. It is incumbent on the author to figure out why we in a different country might care about such issues, and this means more than simply including a glossary of acronyms.

(8) Salami slicing belongs in butcher shops

No one needs to be reminded of the pressure to publish. Every junior faculty member loses sleep wondering if she has enough papers. One way to boost the apparent output is to create



4 G. Norman

multiple publications from the same study. This form of "autoplagiarism" really sits on a continuum, from blatant reproduction of a paper in several journals, which likely occurs rarely since this amounts to direct copyright infringement, to legitimately using a large and complicated study to explore several different questions. The main difficulty lies in the grey zone, where it becomes a matter of judgment whether the second paper in the series really is sufficiently distinct to justify a separate article. The prudent course of action is full disclosure to the editor at the time of submission, so he can judge whether or not to proceed. However, it seems that the prevalence of this problem is on the rise, so AHSE will now require authors to fully disclose any overlap with other submissions.

A corollary to this problem is the "cut and paste" phenomenon, where sections of a paper (e.g., parts of the methodology section) are pasted into a second paper. While not a major crime, this still, strictly speaking, amounts to copyright infringement. The easiest way to avoid it is simply to write the entire paper de novo and not attempt to modify previous text.

(9) If something seems intuitively right, it's likely wrong

One of the neat things about education is that everyone thinks they understand it, since they've had so much personal experience being taught and teaching. But it's almost axiomatic that if something seems self-evidently true, it's likely false. This applies very well to our intuitions about individual differences. We think people are introverted or extraverted, they have visual or verbal learning styles, they are good self-assessors (or not), they are critical thinkers, they have high or low emotional intelligence. However on closer inspection all of these ideas lead nowhere. A number of reviews of learning style come to the same conclusion; it has no educational value (Pashler et al. 2008). People cannot self-assess anything (Eva and Regehr 2005). Self report measures of emotional intelligence have no validity (Lewis et al. 2005). And on it goes.

So do not start a study of learning style, critical thinking ability, emotional intelligence or self-assessment skills. Do not devise an intervention where the outcome is self-assessed knowledge, skills or whatever. We will not publish it; in fact we will not send it out for review. ASHE guidelines are very specific in this regard. Admittedly this issue is simply a subset of point (2) above, but the rule is so frequently violated that it is worth an explicit admonition.

(10) P values tell you whether it's zero but little else

Every research paper tells a story. A well written paper is like a detective story, where by the time the data are all sifted through, Ma Nature reveals the solution to the whodunit. But a poorly written paper is like a poor detective story where, when you find out who is the baddie, you can't make the clues add up. Similarly, a paper where the results are of the form "small groups were significantly better than lectures, p < .0001" tells the reader almost nothing about the data. The "<.0001" says that there's one chance in 10,000 that the difference arose by chance if there really is no difference, but it doesn't even say which direction it's in. It says nothing about the importance of the finding—is it a 2 or 20 % difference? At a minimum, the reader is entitled to see the means, the standard deviations, maybe confidence intervals, maybe effect sizes. But they should be able to *see* what the data are telling them Figures and graphs really are often worth a thousand words.

And some number etiquette, if you please. The average age was not 23.17689 years; it was 23.2 years. The p value was not .04368, it was .04. And it was not .00000; nothing in our game has no possibility of happening. It was <.0001.



(11) Data dredging leads to a lot of mud but very little gold

While we're on the subject, *don't data dredge*. If you have put together a survey with 140 items, do not do a t test on each item—total score only. If you have multiple outcomes, do a Bonferroni correction and set the alpha level at .05/(number of tests). If you have created a table with 132 correlations, do not cull through to find the 5 % that are significant at the p < .05 level, and tell a post hoc story about them.

(12) More research is always required

Sometimes we researchers are as single-mindedly self-serving as the National Rifle Association. Have you ever read an article that concludes with "Well that pretty well put this question to bed. Now we really understand what's going on"? Of course not. We always want more research; if there was no more research we would all be out of a job. If you're going to make this claim, you have to be very specific about precisely what question remains to be answered. Otherwise, it's just self-serving platitudes.

Conclusion

This is a bit different from most of my editorials. While it's written in a flippant style, the substance is deadly serious. Enormous resources are invested in the research enterprise. The "product" of this enterprise is rarely a thing; far more often it is a research paper, and far too many research papers never see the light of day. It is a human tragedy that so many of these efforts are in vain. It is my hope that, in some small way, the community may benefit from some of these observations.

References

- Campbell, D. T., Stanley, J. C., & Gage, N. L. (1963). Experimental and quasi-experimental designs for research (pp. 171–246). Boston: Houghton Mifflin.
- Cook, D. A. (2005). The research we still are not doing: An agenda for the study of computer-based learning. *Academic Medicine*, 80(6), 541–548.
- Cook, D. A., Levinson, A. J., Garside, S., Dupras, D. M., Erwin, P. J., & Montori, V. M. (2008). Internet-based learning in the health professions. *JAMA: The Journal of the American Medical Association*, 300(10), 1181–1196.
- Eva, K. W., & Regehr, G. (2005). Self-assessment in the health professions: A reformulation and research agenda. *Academic Medicine*, 80(10), S46–S54.
- Issenberg, S. B., Petrusa, E. R., McGaghie, W. C., Felner, J. M., Waugh, R. A., Nash, I. S., et al. (1999). Effectiveness of a computer-based system to teach bedside cardiology. *Academic Medicine*, 74(10), S93–S95.
- Lewis, N. J., Rees, C. E., Hudson, J. N., & Bleakley, A. (2005). Emotional intelligence medical education: Measuring the unmeasurable? *Advances in Health Sciences Education*, 10(4), 339–355.
- Norman, G. (2011). Fifty years of medical education research: Waves of migration. *Medical Education*, 45(8), 785–791.
- Norman, G., Dore, K., & Grierson, L. (2012). The minimal relationship between simulation fidelity and transfer of learning. *Medical Education*, 46(7), 636–647.
- Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles concepts and evidence. *Psychological science in the public interest*, 9(3), 105–119.

