Adv in Health Sci Educ (2017) 22:577–580
DOI 10.1007/s10459-017-9780-3

EDITORIAL

Why?

Geoff Norman[1]

I presume, as you read the title of this editorial, the first thing that comes to mind is "Why would anyone write an editorial called 'Why'?" All will be revealed in due course. But first, permit me to describe the personal context that led to this essay.

Very soon, I will appear in a Roundtable session at McMaster, on what has been titled the Geoffrey Norman Educational Research Day. I thought it was a singular honour to have a day named after me until a colleague pointed out that the acronym is G-NERD. Perhaps appropriate. However while I was honored, I was also somewhat anxious. The problem is that I am to appear with Lorelei Lingard. Her name will be known to many of you. In medical education circles, she is on the short list of divas of qualitative research. She singlehandedly dispelled the myth that only test tube rattlers and trialists can get federal funding in health sciences. Sharing a stage with Lorelei was sufficient to cause some lost sleep.

But there were a few other reasons. My initial fear was that everyone expected us to stage the great debate on qualitative versus quantitative research; a debate that promised to make Clinton versus Trump look almost genteel. And that violated two of my lifelong canons: (1) Never engage in any debate. They shed more heat than light. And (2) Never, NEVER debate either a philosopher (my son) or a rhetorician (Lorelei). These folks have spent their lifetime honing their use of language. Us humble number-crunchers are doomed to lose.

My fears were somewhat allayed when Lorelei and I met and agreed that a debate format, and in particular a debate on the old qualitative/quantitative question would be, at best, stultifyingly boring. Instead, we decided on a discussion about "How do you get a good research question? The net effect of this shell game was to replace one anxiety with another. I'm perfectly happy with running workshops on research design, measurement, statistics—all that stuff. There's tons to talk about and lots of resources. Getting answers is

✉ Geoff Norman
norman@mcmaster.ca

[1] McMaster University, Hamilton, ON, Canada

the easy part. But when someone asks me to talk about where do you find a good question, I draw a complete blank.

Anyway, the rubber was to hit the road in a week. What could I say? I know I have come up with some good questions in the past. (Interestingly my perception is that the past few years have been more productive in that regard than the previous decades. No doubt a recency effect). But how did this happen? How, in fact, do we know that it's a good question?

Fortunately, the discussions Lorelei and I had created fertile ground. One thing became obvious from the outset. Although we approach the field with drastically different epistemologies, and as a consequence ask very different questions, at a deeper level, when we reflected on our views on science, we found little to disagree about. The discussion was fruitful, and I'm sleeping much better now.

So how do we think about what constitutes good science? One way is to read a bit of philosophy of science. Popper said that good science amounts to refuting cautious hypotheses, and affirming bold hypotheses. I've done both. I originally was noticed when I did a study of something called a "Patient Management Problem (McGuire and Babbott 1967), a paper problem where the candidate was given long lists of things they could do on history, physical etc. and a pen to rubout the answer box and expose how long the cough had been going on. Each box had a weight from $+2 \rightarrow -2$ and then a score was derived based on how many $+2$'s and $+1$'s you picked and how many -2's and -1's you missed. It was supposed to simulate clinical problem-solving. My question was the essence of simplicity, "Do docs do the same thing with PMPs as they do with the real patient?" We got a bunch of problems, created a PMP version and a standardized patient version, and out some docs through both. We found that people did exactly twice as much with the PMP as they did with the SP (Norman and Feightner 1981). And that was that!

The trouble is that I rapidly got a reputation as an iconoclast (from the Greek—"statue breaker"). And indeed, carry this approach of refuting cautious hypotheses far enough and all you have to show for it is a field of busted marble.

What, then about affirming bold hypotheses? Well, I've been researching clinical reasoning since those days. One constant is an interest in the process of early hypothesis generation. We called it "Non-analytical reasoning"; now it's called "System 1 thinking". In any case, it happens rapidly and effortlessly, usually at the beginning of the encounter (Gruppen et al. 1987). The prevailing wisdom (Croskerry 2003) is that errors are a consequence of sloppy System 1 thinking and would go away if we could get people to slow down, be systematic, gather all the data, etc. I didn't think so. System 1 is too good to be sloughed off that easily So we did a study to look at the relationship between time and accuracy, hypothesizing that accuracy will be associated with lower, not greater time. And it was (Norman et al. 2014).

I think both studies addressed good questions. Of course, I may be just a teeny bit biased. But this begs the issue of how do we know it's a good question. We can't fill out a CONSORT (Schulz et al. 2010) checklist; that tells you how good the methods are, not how good the question is. But one clue comes from listening to how scientists talk about these studies. The ones that capture coffee break time are frequently described with aesthetic terms. They're "cool", "neat", "gorgeous"—even "sexy" (ever seen a sexy F test? Me neither). They're not described as rigorous, tight, unbiased. The review article may talk that way, but then these aren't the attributes that catch our attention. In short, great studies are more likely to be defined by the question they ask than the methods they use to get to an answer. Yes there are some exceptions as in when a research comes up with a really novel approach to design or analysis. But I think these are quite rare. The questions are bold and elegant, and if they're asked carefully enough, the answers are clean and unequivocal and don't require eleven tables and multivariate analysis to prove. As well as elegance, great studies frequently adhere to Occam's razor—clean parsimonious explanations. (On the other hand, good questions frequently lead to further questions as our understanding deepens, a point we will return to forthwith).

So now we have some insight into what a good question looks like. But we still don't have any insight into how to find one. Let me suggest, though, some approaches that don't qualify. To do this, I turn to a wonderful paper by Cook et al. (2008) that described three fundamentally different kinds of research question:

1. Description

   > "I have a new (curriculum, questionnaire, simulation, OSCE method, course) and here's how I developed it"

   That's not even poor research. It's not research at all.
2. Justification

   > "I have a new (curriculum, module, course, software) and it works. Students really like it" OR "students self-reported knowledge was higher" OR "students did better on the final exam than a control group" OR even "students had lower mortality rates after the course"

   OK, it's research. But is it science? After all, what do we know about how the instruction actually works? Do we have to take the whole thing on board lock, stock and barrel, to get the effects? What's the active ingredient? In short, WHY is it better? And that brings us to
3. Clarification

   > "I have a new (curriculum, module, course, software). It contains a number of potentially active ingredients including careful sequencing of concepts, imbedding of concepts in a problem, interleaved practice, and distributed practice. I have conducted a program of research where these factors have been systematically investigated and the effectiveness of each was demonstrated".

That's more like it. We're not asking if it works, we're asking why it works. And the results truly add to our knowledge about effective strategies in education. So one essential characteristic is that the findings are not limited to the particular gizmo under scrutiny in the study. The study adds to our general understanding of the nature of teaching and learning.

Science does not have an accepted definition. It clearly has no single approach—I haven't a clue what "scientific method" would be common to my previous work in nuclear physics, my current work in cognition, and Lorelei's research into team competence and communication. But all have some recognizable common characteristics. They are all striving to address the "why" question, seeking generalizable and universal knowledge based on systematic research. That makes it different not just from religion or journalism; it also makes it different from many applications of scientific methods to unscientific ends. Some of the best and most sophisticated social science and statistical approaches are now being used by people at Ipsos and Google to understand consumer choice. Does it use scientific methods? Unquestionably. Is it science? Unquestionably not.

But we're still left with no real insight as to what a "why" question looks like, other than it begins with "why". Where to from here? One insight arises from my role as editor. When I examine closely the factors that reviewers cite in a decision to reject an article, one frequent comment is "no theoretical framework". I confess I have trouble with that statement. Perhaps because the kinds of theories we advance in trying to understand learning issues seem awfully crude compared to the theories I had to understand in order to pass (barely) exams in physics—quantum mechanics, special relativity, electromagnetic theory, etc. Ours seem much more ad hoc, much more local in application, and much cruder. We're happy if things go in the right direction and make no pretense of being able to estimate the magnitude of the effect.

But maybe a theory is both more and less than that. Maybe it's simply an explicit expression of the factors or variables that underlie a phenomenon derived from research evidence. That can encompass theories associated with any kind of research, qualitative or quantitative. It really addresses the "Why?" question. And the most critical aspect of a theory is that, instead of addressing a simple "Does it work?," to which the answer is "Yup" or "Nope", it permits a critical examination of the effects and interactions of any number of variables that may potentially influence a phenomenon. So, one question leads to another question, and before you know it, we have a research program. Programmatic strategies inevitably lead to much more sophisticated understanding. And along the way, if it's really going well, each study leads to further insights that in turn lead to the next study question. And each answer leads to further insight and explanation.

The nature of science was summarized beautifully by a Stanford professor of science education, Mary Budd Rowe, who said that:

> Science is a special kind of story-telling with no right or wrong answers. Just better and better stories.

That makes a perfect last word.

References

Cook, D. A., Bordage, G., & Schmidt, H. G. (2008). Description, justification and clarification: A framework for classifying the purposes of research in medical education. *Medical Education, 42,* 128–133.

Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine, 78,* 775–780.

Gruppen, L. D., Woolliscroft, J. O., & Wolf, F. M. (1987). The contribution of different components of the clinical encounter in generating and eliminating diagnostic hypotheses. In *Research in medical education: Proceedings of the… annual conference. Conference on research in medical education* (Vol. 27, pp. 242–247).

McGuire, C. H., & Babbott, D. (1967). Simulation technique in the measurement of problem-solving skills. *Journal of Educational Measurement, 4,* 1–10.

Norman, G. R., & Feightner, J. W. (1981). A comparison of behaviour on simulated patients and patient-management problems. *Medical Education*, *15*, 26–32.

Norman, G., Sherbino, J., Dore, K., Wood, T., Young, M., Gaissmaier, W., et al. (2014). The etiology of diagnostic errors: A controlled trial of system 1 versus system 2 reasoning. *Academic Medicine, 89,* 277–284.

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *BMC Medicine, 8*(1), 18.