Specialization and the road to academic success

Peer-reviewed letter

Scientists must choose whether to focus on a few research areas or to apply their skills to a wider range of topics. Their choice may have direct consequences for their academic success. Narrow research breadth allows a scientist to become an authority on a given subject (Hackett 2005) – traditionally a key academic requirement. On the other hand, studying a wide range of topics may enable a researcher to tackle problems that require a broader view, possibly integrating several fields of inquiry. Although the question of whether to specialize or not is encountered by all scientists, few empirical studies have explored how this choice influences success.

In this study, we relate the research breadth of ecologists and evolutionary biologists to a commonly used measure of academic success. We estimate research breadth using ISI Web of Science (ISI; www.isiwebofknowledge.com) standardized keywords, which are independent of author choice of keywords or title. Our premise is that similarity in keywords among publications is related to similarity in the underlying topics studied by an author (Leahey 2007). To quantify this similarity, we use the slope of the log–log transformed cumulative number of unique keywords against the cumulative total number of keywords accumulated per publication throughout an author's career (Figure 1a). Higher slopes indicate that authors are accumulating new keywords between publications at a faster rate, and hence have broader research breadth than authors with lower slopes. Although success is hard to quantify, many institutions use bibliometric indices as a simple proxy for success. We use the H-index (Hirsch 2005) because it integrates two prominent and easily obtainable facets of academic research: the number of publications and the number of citations those publications have attracted. Although the H-index can be influenced by factors that were not controlled for in this study, such as an author's gender or country of residence (Kelly and Jennions 2006), it has strong predictive power (Hirsch 2007). We extracted a list of all publications (limited to the years 1980–2010) for more than 4000 ecologists and evolutionary biologists, along with keywords and number of citations, from the ISI database (WebPanel 1). We then analyzed the relationship between research breadth and H-index for these authors. The H-index is constrained by the number of publications, so even a highly successful early-career researcher is likely to have a low value. To examine how the relationship between the H-index and research breadth changes with number of publications, we sorted authors into publication quartiles (see Figure 1 caption) and repeated the analysis.

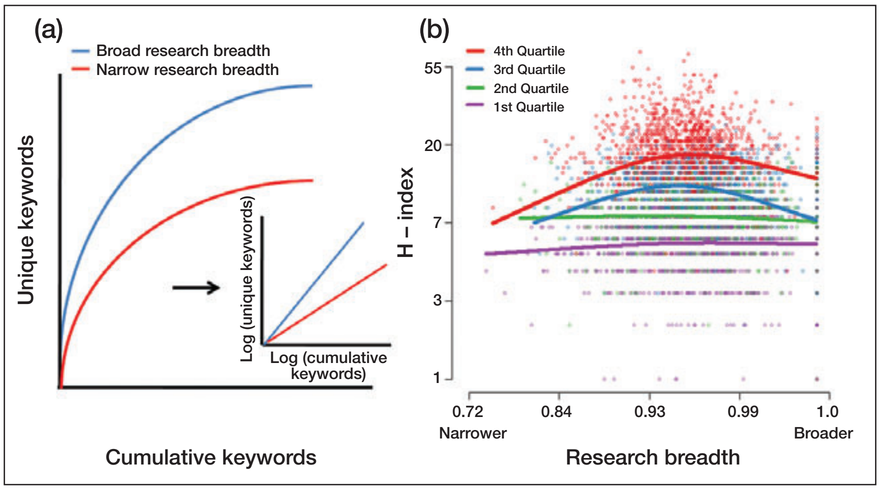

Figure 1. (a) Schematic illustration of research breadth calculation. Research breadth of an author is defined as the slope of the log-transformed cumulative number of unique keywords versus log-transformed cumulative total keywords. (b) The relationship between H-index and research breadth across authors. Separate generalized additive models (GAMs) were constructed for each publication quartile: (1) 10–15 publications (F = 2.80, P = 0.07, % deviance explained = 0.70, n = 1082); (2) 16–24 publications (F = 2.70, P = 0.07, % deviance explained = 0.70, n = 1042); (3) 25–42 publications (F = 86.8, P < 0.01, % deviance explained = 13.3, n = 1085); (4) > 42 publications (F = 82.8, P < 0.01, % deviance explained = 14.2, n = 1122). We modeled GAMs using a smoothing spline with three degrees of freedom and an identity link with Gaussian errors.

We find clear evidence that the relationship between the H-index and research breadth is unimodal, with the highest values achieved at intermediate levels of research breadth (Figure 1b). Our results are strongest for the subset of authors with the most publications (the top two quartiles). It is unclear whether the relationship for authors with fewer publications is absent or is simply harder to detect because of the upper bound on the H-index imposed by few publications.

Our results demonstrate that both over-specialization and over-generalization are detrimental to academic success. Presumably, overly specialized researchers are unable to publish highly influential and widely cited papers because the wider context of their research is not immediately obvious to others. They may also fail to switch research fields when their field becomes obsolete. Conversely, the data also suggest that being a "jack-of-all-trades" really can make you a "master of none", at least according to one's publications and their corresponding citations. Researchers should maintain broad research programs by modifying research focus through time or engaging in out-of-field collaborations (Lee and Bozeman 2005), while ensuring this does not come at the cost of losing familiarity with one's own particular research specialties.

We find that research specialization influences one measure of academic success (H-index). However, we caution that success is not just about publications and citations. The challenge facing every scientist is to decide what really constitutes "academic success" and subsequently to identify the best ways to achieve it.

Jonathan Belmaker[1*], Natalie Cooper[1], Tien Ming Lee[1,2], and Hamish Wilman[1,2]
[1]Yale University, New Haven, CT [*](jonathan.belmaker@yale.edu); [2]University of California, San Diego, La Jolla, CA;
all authors contributed equally

Hackett EJ. 2005. Essential tensions: identity, control, and risk in research. *Soc Stud Sci* **35**: 787–826.
Hirsch JE. 2005. An index to quantify an individual's scientific research output. *P Natl Acad Sci USA* **102**: 16569–72.
Hirsch JE. 2007. Does the *h* index have predictive power? *P Natl Acad Sci USA* **104**: 19193–98.
Kelly CD and Jennions MD. 2006. The h index and career assessment by numbers. *Trends Ecol Evol* **21**: 167–70.
Leahey E. 2007. Not by productivity alone: how visibility and specialization contribute to academic earnings. *Am Sociol Rev* **72**: 533–61.
Lee S and Bozeman B. 2005. The impact of research collaboration on scientific productivity. *Soc Stud Sci* **35**: 673–702.

doi:10.1890/10.WB.25

Statistical danger zone

In a recent review, Ellison and Dennis (*Front Ecol Environ* 2010; **8[7]**: 362–70) argued that advanced statistical fluency is essential for ecologists and identified paths to increase such fluency. We agree that the next generation of ecologists needs a solid understanding of the statistical theory underlying both standard and cutting-edge applications. All ecologists should be able to postulate clear hypotheses, and then – on their own or with assistance – translate alternative hypotheses into probability models tailored for their specific investigation (Hilborn and Mangel 1997). A well-designed series of integrated courses can bring students far beyond the traditional approach of teaching calculus as prerequisite material to be forgotten, statistics as a "grab bag" of analytical tools, and both subjects as disconnected from ecological theory and practice. Such an integrated approach could further serve a fundamental principle for effective instruction: teaching in context. This would improve motivation and help students retain and apply these concepts more effectively (Millenbah and Millspaugh 2003).

However, Ellison and Dennis greatly overestimate how far down the path of statistical fluency the average ecologist will go with a few semesters of probability theory and calculus, no matter how well integrated. Essentially the authors are suggesting that instead of teaching a teenager to drive a car safely, we can tweak driver education so the average 16-year-old becomes able to pilot a fighter jet. Some ecologists will develop sophisticated understanding of statistical theory and implementation during or after graduate school, but usually this will require advanced coursework, collaborative experience working with a quantitative expert, a great deal of self study, and prodigious effort. This minority of ecologists may develop the capability for extending complex statistical methods to novel situations. But we are skeptical that redesigned coursework as described by Ellison and Dennis will offer most other ecologists a shortcut to statistical self-sufficiency.

In fact, we view "self-sufficiency" as a misguided goal. Most ecologists will never develop the statistical expertise of practicing, master's-level statisticians. Ecologists need to recognize the value of investing in statistical help from the start of a study, become fluent in statistical basics and conversant with statisticians, and invest time and money in accessing such help. Furthermore, ecologists' agencies and institutions should facilitate collaboration and greater access to statistical expertise. Society cannot afford for us to view statistical partnerships as simply nice when convenient.

Ellison and Dennis's commentary places little emphasis on the most critical quantitative skill needed by ecologists: designing data collection (ie experimental design and sampling design) that will support reliable inference regardless of whether the resulting data are analyzed with a *t* test or a stochastic differential-equation model. In our view, too many ecologists think that complex analytical methods can extract meaningful results from any dataset, irrespective of how those data were collected. Yet even in a calculus-free environment, students can learn the critical principles involved in designing solid, robust, and meaningful studies, and can learn that these principles, not complex analyses, determine the validity of studies.

Joshua J Millspaugh[*] and Robert A Gitzen
Department of Fisheries and Wildlife Sciences, University of Missouri, Columbia, MO
[*](MillspaughJ@missouri.edu)

Hilborn R and Mangel M. 1997. The ecological detective: confronting models with data. Princeton, NJ: Princeton University Press.
Millenbah KF and Millspaugh JJ. 2003. Using experiential learning in wildlife courses to improve retention, problem solving, and decision-making. *Wildlife Soc B* **31**: 127–37.

doi:10.1890/10.WB.26

A reply to Millspaugh and Gitzen

Current graduate programs in the ecological sciences neither equip aspiring ecologists with sufficient ability to draw reliable conclusions from data collected with non-standard designs or existing and emerging observation systems, nor do they prepare them to understand thoroughly the disciplines' central theories. In response to this deficiency, the course-based solution we prescribed for graduate students in ecological sciences (*Front Ecol Environ* 2010; **8[7]**: 362–70) includes four semester-length courses, amounting to one year of "real" calculus and one year of senior undergraduate proba-

