Predicting Publication Success for Biologists

WILLIAM F. LAURANCE, D. CAROLINA USECHE, SUSAN G. LAURANCE, AND COREY J. A. BRADSHAW

Can one foresee whether young scientists will publish successfully during their careers? For academic biologists on four continents, we evaluated the effects of gender, native language, prestige of the institution at which they received their PhD, the date of their first publication (relative to the year of PhD completion), and their pre-PhD publication record as potential indicators of long-term publication success (10 years post-PhD). Pre-PhD publication success was the strongest correlate of long-term success. Gender, language, and the date of first publication had ancillary roles, with native English speakers, males, and those who published earlier in their career having minor advantages. Once these aspects were accounted for, university prestige had almost no discernable effect. We suggest that early publication success is vital for aspiring young scientists and that one of the easiest ways to identify rising stars is simply to find those who have published early and often.

Keywords: academic performance, gender, language, publication success, university ranking

The academic world can seem obsessed with metrics of a researcher’s performance, of which publication frequency is among the most important (Fischer et al. 2012). Further measures of research impact, such as personal citation rates and journal rankings, are also key indicators of academic success (Symonds 2004, Hirsch 2005, 2007, Acuna et al. 2012).

For researchers, two realities seem unlikely to change in the foreseeable future. First, individual scientists vary greatly in their publication rate (Allison and Stewart 1974). Second, employment opportunities, grant success, and professional accolades are often tied intimately to one’s publication prowess (Zuckerman 1967, Reskin 1977, but see Leahey 2007). Given these realities, we sought to determine whether one can predict who is likely to publish prolifically during his or her scientific career and who is not.

We tested five characteristics that might plausibly be used to gain insight into an academic’s future publication success. The first factor that we assessed was gender. We included this factor because prior studies have suggested that female scientists generally produce fewer publications than do male scientists and are more poorly represented on the upper rungs of the academic ladder (Long 1992, Holt and Webb 2007, Ceci and Williams 2011, Dugdale et al. 2011). Many possible explanations have been considered for this trend, including the heavy demands of motherhood during the crucial early phases of a woman’s career (Monosson 2008, McGuire et al. 2012, O’Brien and Hapgood 2012), potential gender bias (Brown 2008), a tendency for women to avoid self-promotion (Moss-Racusin and Rudman 2010), and feelings of isolation among female researchers (Nature 2011), among other reasons (see also Laurance et al. 2011).

The second factor was whether English was the researcher’s native language. English has become the dominant language of international science; for example, 87% of all papers listed in Biological Abstracts are in English, with no other language constituting over 2% (Monge-Nájera and Nielsen 2005). For those for whom English is not their first language, proficiency varies greatly and often influences their attitude toward publishing in English-language journals (Ferguson et al. 2011). To be competitive for high-ranked journals, those with limited English skills may be forced to collaborate with native English speakers, to use commercial editing services, or to have their key works translated into English (Meneghini and Packer 2007, Primack and Marrs 2008).

A third predictor was the overall prestige of the university from which the researcher received his or her doctorate. This factor could show effects if higher-ranked universities attract better, more-motivated students, who are more likely to succeed in the long term. Potentially contributing to this are greater financial resources for research and a “culture of success” evident at prestigious universities. In most metrics in which universities are compared, the number...
and impact of research publications are strongly weighted (Altbach 2010).

The final two predictors relate to the early publication record of the researcher. The first is the date of the first refereed publication relative to the year of PhD conferral. We predicted that those who published earlier in their careers would, ultimately, be more productive. The second is the total number of refereed papers produced before receipt of the PhD (including papers published the same year the PhD was awarded). We hypothesized that, for whatever reason, some doctoral students simply begin publishing sooner than do others and that this might then carry on throughout their career. As they age, those with early publishing success might also gain more funding and more opportunities for research (Allison and Stewart 1974), such as that provided, for example, by fewer teaching responsibilities.

Here, we evaluate the impact of these five predictors on long-term publication success. Our aims are to provide selection panelists with useful indicators of potential long-term performance and to highlight some strategies that might help early-career researchers become more successful at publication.

A survey of biological and environmental scientists

At the outset, we surveyed 1447 academics at 35 universities in North America (Canada, Mexico, United States), South America (Brazil, Chile, Colombia, Ecuador), Europe (France, Germany, Norway, Portugal, United Kingdom), and Australia. The academics were all in the biological and environmental sciences (i.e., zoology, botany, ecology, evolution, genetics, environmental science and policy). We excluded other fields of science, because they can have different publication rates and patterns of collaboration (Larsen and von Ins 2010), which might distort our findings.

Of the 1447 potential candidates, we identified 182 who met two key criteria for detailed analysis. They (1) had completed their PhD before 2000 (giving us a 10-year window after the PhD to assess publication success) and (2) had an updated copy of their curriculum vitae (CV) available online (i.e., with information on their publication record, as well as data on gender, the year of PhD completion, and the university from which the PhD was granted). For these candidates, we also determined (through their CV, e-mail correspondence, or our personal knowledge) whether English was their first language. We included candidates regardless of their tenure status.

Notably, less than one-third (32.4%) of the researchers in the academic departments that we surveyed were female (range: 11%–62%), and so we attempted to reduce the numerical bias toward male subjects. To do this, we e-mailed 62 female academics whose PhD was granted before 2000 but whose CV was not provided online and requested an updated copy (prior studies have shown that female academics tend to have a personal homepage, which is often a source for an online CV, less frequently than do male academics; Barjak 2006). Of those who responded, 21 fit our selection criteria and were included in our sample of 182 subjects.

The universities that we surveyed varied widely in their ratings of prestige (with rankings ranging from 1 for Harvard University to higher than 1200 for some universities). The rankings were based on the Academic Ranking of World Universities (ARWU; www.arwu.org). This index has been lauded for being consistent and transparent and for relying solely on independent, objective indicators, not reports generated by the universities themselves (Altbach 2010).

The response variable in our analysis—our measure of academic success—was the number of refereed papers a researcher had published in the decade immediately following the year of PhD conferral. We included only peer-reviewed papers in journals listed in the Web of Science, regardless of whether the researcher was the lead author. Of course, our response variable does not include other measures of scientific success, such as the number of citations a researcher receives. However, we found that our response variable was strongly and linearly related (Pearson’s correlation coefficient, r(117) = .828, p < .00001) to the h-index for each researcher, calculated 10 years after their PhD was completed. The h-index (Hirsch 2005) incorporates data on both citations and publications but was not consistently available for all researchers in our study (this is because Elsevier’s Scopus database, which we used to generate h-indexes, does not incorporate papers published prior to 1996). From this comparison, we conclude that our simple response variable provides a reasonably robust measure of one’s publication success as measured by the number of citations.

Because publication rates can differ among scientific fields, we limited our study to biologists (excluding those in biomedical fields) and environmental scientists. For a subset of our researchers, we also determined their subdiscipline (e.g., ecology or evolution, plant sciences, environmental toxicology). We then contrasted the distribution of subdisciplines between the top 10% and the bottom 10% of subjects in terms of the number of publications in the first decade after PhD conferral. We found little difference overall (e.g., ecology or evolution predominated in the two groups), which suggests that any productivity differences among the subdisciplines were modest in our sample and unlikely to confound our results.

To analyze our data, we applied a generalized linear-model format with a gamma error distribution and log-link function to account for the nonnormal nature of our response variable and for predictor heteroscedasticity. We compared and ranked models using the Bayesian information criterion, which measures their weight of evidence relative to other models and better distinguishes main from tapering effects than does Akaike’s information criterion (Burnham and Anderson 2002, Link and Barker 2006). We assessed each model’s relative probability using Bayesian information...
criterion weights and its structural goodness of fit using the percentage of deviance explained by the model. We also calculated model-averaged standardized coefficients for each response variable and model term. These coefficients show the relative effect of each predictor when the other predictors are controlled statistically, while simultaneously accounting for model uncertainty (Bradshaw et al. 2012). We assessed the predictive ability of the top-ranked model using a 10-fold cross-validation (Davison and Hinkley 1997). This bootstrap-resampling procedure (here, using 1000 iterations) estimates the mean model-prediction error for 10% of observations randomly omitted from the calibration data set.

Key findings
Our sample of established academics included 113 male and 69 female subjects. Over 60% of those in our sample (116) were native English speakers.

Spearman rank correlations (table 1) revealed relatively weak associations among our potential predictor variables, with two exceptions. First, those who published earlier in their career typically had more publications when they completed their PhD (figure 1). Second, there was a weak tendency for native English speakers to attend more prestigious universities than did those with other first languages.

According to the generalized linear models (table 2), the number of publications at the completion of the PhD was the best overall correlate of long-term publication success among those in our sample (116) were native English speakers.

Using the 10-fold cross-validation, our top-ranked model (table 2) had a mean prediction error of 88%, and a cross-validated $R^2$ value of 0.14, which was comparable to the percentage of deviance explained.

Many individuals applying for university jobs have postdoctoral research experience. When we incorporated the effects of postdoctoral productivity—by replacing the number of publications at PhD conferral with the total number of papers published by 3 years after the PhD was completed—the explanatory power of our models improved markedly (the percentage of deviance explained increased to 29.0%–34.4%; table 3). When the postdoctoral years were included, more-productive individuals showed a striking advantage in their long-term publication success, with the other four predictors declining in relative importance (figure 2b).

Our models suggest that native English speakers and male academics enjoy some modest advantages in long-term publication success, but there was great variation among the individual subjects (figure 3). When language and gender were considered as the sole predictors of long-term success (all papers published 10 years after the PhD plus pre-PhD papers), the gender-only model had nearly three times more support (3.4% of the deviance was explained) than did the language-only model (2.3%), based on the evidence ratio.

| Table 1. Spearman rank correlations among five potential predictors in our analysis. |
|-----------------------------------------|----------------|----------------|
|                                        | Publications before PhD | Year of first publication relative to PhD conferral | Gender | Language |
| Year of first publication relative to PhD conferral | −.781 |                |
| Gender                                  | −.128 | .068           |
| Language                                | .041  | .028           |
| University rank                         | −.026 | .034           |
|                                        | −.154 | .248           |

Note: The values in bold indicate evidence for a relationship based on a Bonferroni-corrected p-value ($\rho > .207, p < .005$). Correlations involving the two binary variables, gender and language, should be interpreted cautiously.
There was no evidence for a gender × language interaction (table 4).

In our models, university prestige was a surprisingly weak predictor of publication success (table 2, figure 2). This relationship held even when we used two other leading university-ranking systems in place of the ARWU system (see the supplemental material, available online at http://dx.doi.org/10.1525/bio.2013.63.10.9). University ranking had no association with the number of pre-PhD publications (Spearman rank correlation coefficient ($\rho = -0.066$, $p = .93$) but showed a weak association with the number of publications produced in the 10 years after PhD completion ($\rho = -0.146$, $p = .049$), with graduates of more-prestigious universities having slightly better productivity. Over these intervals combined, graduates of top-100-ranked universities averaged 26.1 papers (standard deviation = 15.4, $n = 76$), whereas graduates of universities ranked outside the top 100 published an average of 23.4 papers (standard deviation = 20.1, $n = 106$).

Conclusions
A caveat of our study is that, by necessity, only those with long-term academic positions were sampled. Obviously, many do not achieve this milestone. This inescapable reality

### Table 2. Top-ranked generalized linear models testing five potential predictors of publication success for 182 academic biologists across four continents.

<table>
<thead>
<tr>
<th>Model</th>
<th>Log-likelihood</th>
<th>Number of parameters</th>
<th>Change relative to the top-ranked model</th>
<th>Weight</th>
<th>Percentage of the deviance explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>~Language+Pubs$_{PhD}$</td>
<td>-711.42</td>
<td>3</td>
<td>0.00</td>
<td>.383</td>
<td>13.2</td>
</tr>
<tr>
<td>~Gender+Language+Pubs$_{PhD}$</td>
<td>-709.18</td>
<td>4</td>
<td>0.73</td>
<td>.267</td>
<td>15.2</td>
</tr>
<tr>
<td>~Language+Pubyear+Pubs$_{PhD}$</td>
<td>-709.92</td>
<td>4</td>
<td>2.21</td>
<td>.127</td>
<td>14.6</td>
</tr>
<tr>
<td>~Gender+Language+Pubyear+Pubs$_{PhD}$</td>
<td>-707.69</td>
<td>5</td>
<td>2.95</td>
<td>.088</td>
<td>16.5</td>
</tr>
<tr>
<td>~Gender+Pubs$_{PhD}$</td>
<td>-713.64</td>
<td>3</td>
<td>4.46</td>
<td>.041</td>
<td>11.2</td>
</tr>
<tr>
<td>~Pubs$_{PhD}$</td>
<td>-716.73</td>
<td>2</td>
<td>5.43</td>
<td>.025</td>
<td>8.4</td>
</tr>
<tr>
<td>~Gender+Language+Unirank+Pubs$_{PhD}$</td>
<td>-709.08</td>
<td>5</td>
<td>5.74</td>
<td>.022</td>
<td>15.3</td>
</tr>
</tbody>
</table>

Note: The tildes (~) imply a correlative relationship between the predictors and the response variable. The plus signs (+) represent additive (as opposed to multiplicative) terms in the model.

**Abbreviations:** Pubs$_{PhD}$, the number of refereed publications by the year PhD was awarded; Pubyear, the year of first publication relative to the year the PhD was awarded; Unirank, the rank of the university from which the PhD was awarded.

---

**Figure 2.** Absolute standardized coefficients showing the relative effects of five variables on long-term publication success for 182 academic biologists. (a) Journal papers published in or before the year of PhD conferral (Pubs$_{PhD}$) are used as a potential predictor. (b) Papers in the first 3 years after the PhD are included (Pubs$_{PhD+3}$). The response variable in panel (a) is the number of papers published in years 1–10 after the PhD, whereas in panel (b), it is the number of papers published in years 4–10 after the PhD. Abbreviations: Firstpub, the year of the first publication relative to the year of PhD conferral; Unirank, the ranking of the university from which the PhD was awarded.
influenced our study design, forcing us to modify our selection criteria to include more women academics to help balance our sample sizes. It also constrains our conclusions: If one included individuals who, for whatever reason, had dropped out of academia, the predictors of failure and success might well be different (see also Holt and Webb 2007, Dugdale et al. 2011). In essence, our results apply only to people who stay in the academic game.

Nonetheless, among the 182 established academics whom we surveyed, there was great variation in publication success, ranging from 0 to 87 publications in the first decade after PhD completion. Our findings suggest four conclusions, based on the variables we evaluated. First and foremost, although there was considerable unexplained variation in our models, the best correlate of long-term publication success was early publication success. This trend is most striking when the first 3 years of postdoctoral publications were included along with pre-PhD publications (figure 2b) but holds at the PhD level as well (figure 2a). If one is comparing different candidates for academic or research jobs, simply tallying the number of early-career publications (at some standardized point in one’s career, such as 3 years post-PhD) appears to be an effective way to identify prospective rising stars. Obviously, one would consider a range of additional factors in hiring decisions (Lee and Bozeman 2005, Leahey 2007), but a proclivity for publishing is often vital in today’s competitive academic market.

Second, those who publish earlier in their career—at least a few years before PhD completion—have a clear advantage, because this places them on the fast track for longer-term publication success (figure 1; see also Allison and Stewart 1974, Primack and Stacy 1997). Collectively, our first two findings highlight a crucial role for early training and mentorship for aspiring academics. We assert that the best way to promote the long-term success of one’s graduate students is to assist them in publishing early and establishing this as a key performance indicator for both students and their graduate supervisors. Additional goals, such as publishing in high-impact journals (Symonds 2004) and collaborating frequently (Lee and Bozeman 2005), can also enhance one’s long-term scientific impact. On the basis of our personal experience and that of others (e.g., Richard B. Primack, Department of Biology, Boston University, personal communication, 25 October 2012), young scientists who are the lead author on one or more high-impact papers are often advantaged when seeking academic jobs.

Third, language and gender appear to contribute to one’s research success, with male academics and native English speakers having a modest advantage (figure 3). Language and gender would probably have been more important had we been able to include in our analysis those who fail to

### Table 3. Top-ranked predictors of long-term publication success for 182 academic biologists when one adds 3 years of postdoctoral publications to those produced up to PhD completion.

<table>
<thead>
<tr>
<th>Model</th>
<th>Log-likelihood</th>
<th>Number of parameters</th>
<th>Change relative to the top-ranked model</th>
<th>Weight</th>
<th>Percentage of the deviance explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>~Language+Pubyear+Pubs$_{PhD+3}$</td>
<td>~639.05</td>
<td>4</td>
<td>0.00</td>
<td>.389</td>
<td>34.1</td>
</tr>
<tr>
<td>~Language+Pubs$_{PhD+3}$</td>
<td>~642.09</td>
<td>3</td>
<td>0.86</td>
<td>.253</td>
<td>32.0</td>
</tr>
<tr>
<td>~Pubyear+Pubs$_{PhD+3}$</td>
<td>~642.83</td>
<td>3</td>
<td>2.36</td>
<td>.120</td>
<td>31.5</td>
</tr>
<tr>
<td>~Pubs$_{PhD+3}$</td>
<td>~646.30</td>
<td>2</td>
<td>4.08</td>
<td>.051</td>
<td>29.0</td>
</tr>
<tr>
<td>~Gender+Language+Pubyear+Pubs$_{PhD+3}$</td>
<td>~638.56</td>
<td>5</td>
<td>4.23</td>
<td>.047</td>
<td>34.4</td>
</tr>
<tr>
<td>~Gender+Language+Pubs$_{PhD+3}$</td>
<td>~641.40</td>
<td>4</td>
<td>4.69</td>
<td>.037</td>
<td>32.5</td>
</tr>
<tr>
<td>~Language+Pubyear+Unirank+Pubs$_{PhD+3}$</td>
<td>~638.82</td>
<td>5</td>
<td>4.73</td>
<td>.037</td>
<td>34.2</td>
</tr>
</tbody>
</table>

Note: The response variable is the number of refereed papers produced in years 4–10 after PhD completion. The tildes (~) imply a correlative relationship between the predictors and the response variable. The plus signs (+) represent additive (as opposed to multiplicative) terms in the model.

Abbreviations: Pubs$_{PhD+3}$, the number of refereed publications by three years after the PhD was awarded; Pubyear, the year of first publication relative to the year the PhD was awarded; Unirank, the rank of the university from which the PhD was awarded.
attain or who elect not to pursue long-term academic positions. For instance, depending on the academic field and nation, the proportion of female scientists declines markedly as one moves up the academic ladder, from 40%–77% of those receiving a PhD to around 10% of full professors (Holt and Webb 2007, Ceci and Williams 2011). This striking disparity highlights a strong filtering effect that occurs in contemporary academia (Primack and O’Leary 1993, Dugdale et al. 2011). We suspect that varying English-language ability creates a similar filter (Bornmann and Mungra 2011), given that those with other first languages vary considerably in their English writing skills (Ferguson et al. 2011) and their acceptance rates in peer-reviewed journals (Primack and Marrs 2008, Primack et al. 2009).

Finally, we found a surprisingly weak role for university prestige once the effects of other predictors were accounted for statistically (figure 2). This finding held even when we used two other leading university-ranking systems (see the supplemental material), which indicates that it was not merely an artifact of the ranking system that we employed (i.e., the Academic Ranking of World Universities). One possibility is that individual mentors or lab environments vary widely and are more important than is university reputation in determining long-term publication success. Alternatively, key personal attributes, such as motivation, might vary so widely among individuals that they simply swamp the effects of university ranking. Whatever the explanation, our findings suggest that, if two job candidates in biology have comparable publication records, there would be little justification for automatically favoring the candidate from the more-prestigious university.

In summary, we readily acknowledge that publication success is only one of many factors that can determine a biologist’s advancement and career trajectory. Many other attributes—effectively teaching and mentoring students, forging good relationships with colleagues, developing a strong knowledge base in science—are also important. Nonetheless, we believe that a capacity to design, execute, and publish high-level research is among the most vital of academic skill sets and that motivation and achievements in this area can bring manifold other benefits. We suggest that instilling a desire to publish effectively is one of the best ways to help young scientists advance professionally and that early publication success is a key predictor of long-term research achievement.

Acknowledgments
The Australian Research Council supported this study through an Australian Laureate Fellowship to WFL and a Future Fellowship to CJAB. We thank Hannah Dugdale, Richard Primack, and three anonymous referees for detailed comments on the manuscript.

References cited


Table 4. Top-ranked predictors of total publication output for 182 academic biologists considering only gender and language factors.

<table>
<thead>
<tr>
<th>Model</th>
<th>Log-likelihood</th>
<th>Number of parameters</th>
<th>Change relative to the top-ranked model</th>
<th>Weight</th>
<th>Percentage of the deviance explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>−Gender</td>
<td>−740.70</td>
<td>2</td>
<td>0.00</td>
<td>.445</td>
<td>3.4</td>
</tr>
<tr>
<td>Intercept-only model</td>
<td>−744.04</td>
<td>1</td>
<td>1.49</td>
<td>.212</td>
<td>—</td>
</tr>
<tr>
<td>−Gender+Language</td>
<td>−739.01</td>
<td>3</td>
<td>1.84</td>
<td>.177</td>
<td>5.0</td>
</tr>
<tr>
<td>−Language</td>
<td>−741.78</td>
<td>2</td>
<td>2.17</td>
<td>.150</td>
<td>2.3</td>
</tr>
<tr>
<td>−Gender+Language+Gender×Language</td>
<td>−738.81</td>
<td>4</td>
<td>6.65</td>
<td>.016</td>
<td>5.2</td>
</tr>
</tbody>
</table>

Note: The tildes (~) imply a correlative relationship between the predictors and the response variable. The plus signs (+) represent additive (as opposed to multiplicative [×]) terms in the model.
Larsen PO, von Ins M. 2010. The rate of growth in scientific publication and the decline in coverage provided by the Science Citation Index. Scientometrics 84: 575–603.

William F. Laurance (willlaurance@gmail.com) is a distinguished research professor and Australian Laureate at James Cook University in Cairns, Queensland, Australia; he also holds the Prince Bernhard Chair in International Nature Conservation at Utrecht University, in the Netherlands. D. Carolina Useche is a biodiversity and climate change researcher at the Alexander von Humboldt Institute on Biological Resources Colombia in Bogotá, Colombia. Susan G. Laurance is a tropical leader and senior lecturer at James Cook University, and president of the Association for Tropical Biology and Conservation. Corey J. A. Bradshaw is a professor and future fellow at the University of Adelaide, in Australia.