Ranking Doctoral Programs by Placement: A New Method

Benjamin M. Schmidt, *Princeton University*
Matthew M. Chingos, *Harvard University*

While many bemoan the increasingly large role rankings play in American higher education, their prominence and importance are indisputable. Such rankings have many different audiences, ranging from prospective undergraduates or graduate students, to foundations and government funders, to university administrators identifying strengths and weaknesses of their school. This diverse audience necessarily has varying hopes for what "quality" is measured in school rankings, and different uses for the rankings themselves. But although there are currently a wide variety of ways to assess graduate school quality, most existing surveys have recognized failings that compromise their usefulness to at least one of these different constituencies.

Traditionally, the most influential rankings systems have relied on surveys that ask prominent members of the field to give their assessments of graduate programs.[1] However, these reputational surveys are subject to a number of criticisms. Rankers are asked to evaluate over a hundred programs in their field, which forces them to make many judgments with little prior knowledge or strong experience.[2] Universities with strong programs across many fields will rarely be ranked very lowly on even their lowest quality programs; conversely, less prominent institutions have trouble getting full recognition for the accomplishments of their best programs. Additionally, one or two extremely well-known faculty within a program could lead to it being ranked disproportionately high. Finally, and perhaps most importantly, reputational rankings cannot distinguish between larger and smaller programs. Therefore, larger programs will be systematically advantaged since they have larger numbers of faculty, and thus a greater chance of recognizability from the perspective of reviewers. But increasingly, the recognized flaws in such reputational surveys have led to attempts to come up with alternative methods. While *US News and World Report* continues to publish a widely read set of rankings that relies only on reputational surveys, the National Research Council (NRC) has in more recent years moved toward including specific statistical material about departments, in addition to continuing to release its respected but infrequent reputational survey.

The rankings that have most often incorporated objective measures of quality are those that have focused on the publications of departmental faculty. The most basic of these studies measure raw faculty publishing output. One recent such study in political science (Hix 2004) ranked schools based on the number of articles their faculty have published in leading political science journals, and included a per-capita measure that corrects for the tendency, in both reputational and output studies, for overrepresentation by large departments. Citation studies, which employ a slightly more complicated measure, judge institutional quality not simply by the number of faculty publications, but by the impact of those publications on the field as measured by citation count. The most recent NRC rankings included both publication and citation counts. But however well founded, departmental research output represents only an incomplete picture of departmental quality. Such studies regard departments entirely as loci of research, not of teaching, and are easily subject to distortion by one or two "star" faculty members who may play an insignificant role in departmental life.

Finally, a focus on publications and citations as the primary tokens of academic achievement can be problematic in a field as diverse as political science. While the journal article is easily distinguishable as the currency of ranking in the natural sciences and in many social science subfields, monographs play the central role in the scholarly discourse in more humanistic disciplines. In political science, this means a focus on journal articles and citations will lead to an artificial advantage for schools that focus more heavily in quantitative fields that publish frequently in journals (for instance, political economy) over schools with strengths in more humanistic areas (such as political theory).

This paper proposes a new ranking method that is based on the presumption that for many users of rankings the new scholars produced by a department may be as important an indicator of departmental quality as the new research produced within it. In this broad philosophy, we follow others (Laband 1986; McCormick and Bernick 1982) who have published rankings based on the research productivity of graduates from Ph.D. programs. But such studies are quite difficult to execute (probably the primary reason they are rarely performed), and like faculty-oriented output rankings, they focus only on a small number of peer-reviewed journals. By focusing instead on graduate programs' history of placing their graduates into faculty positions, we create a metric that allows a less arbitrary means of selection by focusing on a single, crucially important subset of the American academic universe: universities with Ph.D. programs. The reciprocal patterns of faculty hiring among the 100 or so American universities that grant doctoral degrees in political science describe an implicit hierarchy. Our method reveals this hierarchy using techniques already developed for citation analysis.[3]

We have chosen to base our rankings on graduate placement, however, not only for elegance of calculation. One of the great problems with the current landscape of graduate rankings is a "one size fits all" approach to the problem of ranking—somewhat crudely put, an idea that institutional "quality" is a one-dimensional aspect. We aim to present a ranking that, by limiting its scope to one sharply defined facet of quality, allows a

Benjamin M. Schmidt *is a Ph.D. candidate in the department of history at Princeton University and a former staff member of the Humanities Data Initiative at the American Academy of Arts and Sciences.*

Matthew M. Chingos *is a Ph.D. candidate in the department of government at Harvard University and a research associate at the Andrew W. Mellon Foundation.*

 DOI: 10.1017/S1049096507070771

more precise understanding of quality that nonetheless corresponds reasonably well with comprehensive ranking systems. The placement rate actually measures two distinct facets of a graduate program, as it is influenced by both the quality of students that a program is able to attract, and by the value that a program adds to those students over the course of educating them and helping them find employment. Our measure of program quality cannot distinguish between these two factors. However, since both are important factors in education and quite difficult to measure independently, we believe that it is not critically important to separate their roles. This system effectively measures the real-world prominence of an institution's graduates, regardless of the reasons for their prominence, and uses data that are, for most fields, already available. Lastly, it should be noted that this is a lagging indicator: it does not compensate for dramatic increases or decreases in quality toward the end of the time period studied.

Accepting these limitations, we believe, leaves our method with several advantages. Like most measures based on real-world data, this method is not subject to individual biases or sampling and response-rate problems among survey respondents, and it allows us to create both unweighted and per-capita measures. But unlike the previously published ranking methods already discussed, our method is less tied to the fluctuations of the most prominent or productive faculty members, and more focused on the quality of students and graduate education. Moreover, unlike faculty output studies, our measure correlates highly with the subjective rankings, meaning that the results align reasonably well with popular opinion. In addition, our per-capita measure is especially useful to prospective graduate students because placement rates are chiefly determined by two factors of great importance to them—the quality of the peer group of students who matriculate at the program, and the effectiveness of programs in securing prestigious jobs for their graduates.

A New Ranking Method

Raw Placement Rates, and the Suitability of Placement Data for Rankings

Our method takes as its starting point the number of a given program's graduates that have found employment on the faculty of American political science departments. In its rawest form, such data take the form of one simple statistic: the percentage of Ph.D. recipients a program has placed in faculty positions in recent years. These raw numbers are not particularly useful, though, since they do not reflect differences in the quality of the institutions hiring the recent Ph.D. recipients. Thus, two programs that each placed 50% of their respective graduates in tenure-track positions would be ranked equally, even if one program placed its graduates at better universities.

Before explaining the means by which we derive a more refined departmental hierarchy from placement data, it is important to acknowledge some possible questions as to the validity of using placement data for rankings at all. It could be argued that using teaching positions as a stand-in for overall program quality is inadequate, since academic careers are not the only option available to newly minted Ph.D.s. Indeed, much attention in graduate education in the humanities and humanistic social sciences has been paid to preparing graduates for careers outside of academia. Recent studies have found that as few as 55% of graduates of political science doctoral programs have tenured or tenure-track positions 10 years after receiving their Ph.D.s, and only 14% have tenured positions in Research I universities (Nerad and Cerny 2003, 6).[4] Ranking programs on the basis of this fraction of the graduate population might seem of limited use.

On some level, this is true—our metrics, and the rankings they create, are inapplicable for programs in which training for an academic career is not the predominant goal, such as for programs in the sciences. However, recent data indicate that graduate students in many social science and humanities fields (including political science) primarily aspire to academic careers.

The Survey of Earned Doctorates asks graduating Ph.D.s their career plans; recent data show that Ph.D. recipients overwhelmingly pursue academic careers: 77.3% of all 2004 political science graduates with definite plans were continuing on to some sort of academic position, whether teaching or further study, as opposed to just 17.9% choosing any form of business, government, or nonprofit sector (BGN) work. Nerad and Cerny (2003) asked political science Ph.D.s about a decade after graduation what their career aspirations had been upon completion of their degree. They found that 72% of graduates aimed to be professors at the end of their education, with 7% desiring other academic positions and just 11% preferring non-academic research or administration jobs.[5] Overall employment data show that, despite increased discussion of non-academic careers for Ph.D.s, the percentage of Ph.D.s that finds employment in academia has not changed significantly in the last 20 years. In 1995, the last year for which data are available, fully 80% of all humanities Ph.D.s were working in academic institutions (Ingram and Brown 1997). The percentage of humanities Ph.D.s going on to BGN employment actually fell, from 20.8% to 14.0%, between 1984 and 2004 (Hoffer et al. 2005); in the social sciences, the corresponding numbers decreased from 45.8% to 27.6%, largely (but not entirely) due to an increase in postdoctoral study.

Although our method does not count those placed into positions at prestigious liberal arts colleges, the number of positions at such schools is limited compared to those at doctoral institutions. While our method also excludes graduates in other sectors of academic employment (comprehensive universities, two-year colleges, and high schools), it seems unlikely that a great number of those able to find tenure-track employment in doctoral universities would choose such positions instead.

No single ranking can address all the many roles played by graduate education in the humanities and social sciences, but we believe our ranking provides a good objective correlate to the issues of academic quality addressed by many of the existing ranking systems. In addition, its limitations are explicit, unlike those in reputational or publication surveys, so it should be more clear where its application is appropriate—programs that do not view training professors for positions in doctoral universities as a primary goal should not be slighted by falling low on this particular measure.

The Basic Formula

As a starting point for an objective ranking of graduate programs, we take a ranking system that has had success in another application: the "PageRank" formula used by the Google search engine to rank web pages. PageRank looks at the pattern of links on the Internet to see which pages are the most prominent; our application here looks at the pattern of faculty hires to draw out the implicit hierarchy of academic programs (Page et al. 1998).[6]

The only information needed for the ranking is a square matrix with rows and columns corresponding to the schools being ranked. Each row and column will correspond to a school, with the matrix resembling the following:

$$\begin{bmatrix} \ell(p_1,p_1) & \ell(p_1,p_2) & \cdots & \ell(p_1,p_N) \\ \ell(p_2,p_1) & \ddots & & \\ \vdots & & \ell(p_i,p_j) & \\ \ell(p_N,p_1) & & & \ell(p_N,p_N) \end{bmatrix}$$

where $\ell(p_1,p_1)$ is the number of graduates from the first program over a designated time period who hold tenure-track positions at the first program at the time of the ranking, $\ell(p_1,p_2)$ is the number of graduates from the first program who hold tenure-track positions at the second program, and so forth.

This information alone would be sufficient for a very raw ranking of programs simply by the number of graduates that they have placed (as a measure of prominence) and, with easily accessible data on the number of graduates from programs, by their placement ratio (as a measure of prominence per Ph.D. awarded). The rankings we will produce are essentially refined versions of each of these statistics that unveil the implicit hierarchy in hiring patterns. In the first, and simpler, case, it is clear that while an initial ranking of raw numbers of placed graduates gives a rough estimate of prominence, a better measure would take into account the prominence of schools at which graduates are placed.

One can think of the process we use as an election in which graduate programs "vote" for other programs (as well as themselves) by hiring their faculty. The votes of better programs (as defined by this formula) are counted more heavily. One can think of each program as initially having the same weight (an initial vector with entries that are all the same), but these ranks change in subsequent "rounds" of voting. In each round, the scores are summed for each school to produce a new ranking vector, which eventually stabilizes. Thus, we move from the raw data on placement numbers to a more complicated, hierarchical ranking of schools.

Using matrix multiplication, we can represent this process by the following formula, where **R** is the ranking of the schools:

$$\mathbf{R} = \begin{bmatrix} \ell(p_1,p_1) & \ell(p_1,p_2) & \cdots & \ell(p_1,p_N) \\ \ell(p_2,p_1) & \ddots & & \\ \vdots & & \ell(p_i,p_j) & \\ \ell(p_N,p_1) & & & \ell(p_N,p_N) \end{bmatrix} \mathbf{R} \tag{1}$$

The result of this will be a vector of the scores of every program from 1 through N, the number of programs:

$$\mathbf{R} = \begin{bmatrix} \text{Score}(p_1) \\ \text{Score}(p_2) \\ \vdots \\ \text{Score}(p_N) \end{bmatrix} \tag{2}$$

We make the matrix stochastic by normalizing the columns so they sum to one, which eliminates any additional "voting strength" conferred on institutions with larger faculties. However, this does not correct for differences in the size of each institution's program in terms of student graduates, which will likely cause larger programs to be ranked higher only because they have more graduates in the academic job market. This may be desirable in some cases—for prominence measures, it is proper that larger programs score higher—but our preferred measure corrects the formula for size (see the Per-Capita Placement Success section).

With one more correction, the formula is complete. Since the score for any school depends on the score of every program at which it has placed graduates, the score will eventually approach zero at schools that only placed graduates at programs without a placement record. In practice, this would give a number of programs a score of zero despite their having placed at least one graduate, since the schools at which they placed graduates did not themselves place any. This can be corrected with the addition of a constant, q, which is set between zero and one, and represents a baseline score divided across all the schools so that no program that placed at least one graduate ever reaches a score of zero.[7] This allows a final version of the formula, adding q and dividing the base matrix by F_j, the number of faculty at the school:

$$\mathbf{R} = q \begin{bmatrix} \frac{1}{N} \\ \frac{1}{N} \\ \vdots \\ \frac{1}{N} \end{bmatrix} + (1-q) \begin{bmatrix} \frac{\ell(p_1,p_1)}{F_1} & \frac{\ell(p_1,p_2)}{F_2} & \cdots & \frac{\ell(p_1,p_N)}{F_N} \\ \frac{\ell(p_2,p_1)}{F_1} & \ddots & & \\ \vdots & & \frac{\ell(p_i,p_j)}{F_j} & \\ \frac{\ell(p_N,p_1)}{F_1} & & & \frac{\ell(p_N,p_N)}{F_N} \end{bmatrix} \mathbf{R} \tag{3}$$

If q is equal to 0, the scores are simply equal to the dominant eigenvector of the matrix of appointments; as q approaches 1, the rankings of the schools all converge closer together. $q = 0$ produces the most elegant formula and the widest discrimination, but we believe it produces better results to use a slightly larger value, around $q = 0.1$, to ensure that schools that have placed graduates at only the lowest tier still score higher than schools that have not placed any graduates at all.

Although the algorithm may seem somewhat abstract, there is a real-world interpretation of the rankings derived. We earlier described it as a successive set of rounds of voting. Another way of thinking of it is as a process of tracing academic influence. The

process would start by selecting a school at random; then selecting a random professor at that school; and then seeing where he or she went to graduate school. One would then repeat the process on that graduate program, and so forth an indefinite number of times. The final score for each school represents the relative chance that the process has landed on it at any given moment. The constant q introduces a random variable into this factor: at each selection of a school, there is a random chance (one in 10, with $q = 0.1$) that a new school will be selected at random instead of continuing the chain.

Per-Capita Placement Success

While the unweighted influence ranking derived earlier is useful, a per-capita measure has clear benefits in not privileging large programs based on their size alone. It is possible to create such a measure by making one small additional change to the formula. Earlier, the columns were normalized so that each summed to one, which ensured that no school would get a disproportionate vote based on the size of its faculty. Now, we add another step before that, dividing each row by the number of Ph.D.s granted by the institution corresponding to the row. This does not cause the rows to sum to one, as not all Ph.D. graduates find jobs at doctoral universities. It does, however, increase the weight of the smaller programs inversely proportionate to the size of their graduate pool, allowing us to get a measure of program prominence independent of size.[8] The formula thus becomes:

$$\mathbf{R} = q\begin{bmatrix} \frac{1}{N} \\ \frac{1}{N} \\ \vdots \\ \frac{1}{N} \end{bmatrix} + (1-q)\begin{bmatrix} \frac{\ell(p_1,p_1)}{G_1H_1} & \frac{\ell(p_1,p_2)}{G_1H_2} & \cdots & \frac{\ell(p_1,p_N)}{G_1H_N} \\ \frac{\ell(p_2,p_1)}{G_2H_1} & \ddots & & \\ \vdots & & \frac{\ell(p_i,p_j)}{G_iH_j} & \\ \frac{\ell(p_N,p_1)}{G_NH_1} & & & \frac{\ell(p_N,p_N)}{G_NH_N} \end{bmatrix}\mathbf{R} \tag{4}$$

with G_i being equal to the number of graduates of school i over the time period being studied, and H_j replacing the number of faculty in ensuring that the columns sum to one. For each column, it is the sum of the adjusted faculty weights:

$$H_j = \sum_{i=1} \frac{\ell(p_i,p_j)}{G_i} \tag{5}$$

It should be noted that since the final weights are different in this ranking, the final results are not the same as a simple per-capita measure of the unweighted influence score. While such a measure would give a different sort of weighted score, we find that this method produces results that are less likely to vary when influenced by minor changes, particularly in the case of a small, middle-tier school that places just one faculty member in a very good (top five or so) institution.

Ranking Political Science Programs

The most difficult task in almost any ranking system is collecting high-quality data. Ideally, information collected would track the school at which graduates are placed three to five years after they leave graduate school (any earlier would not give careers sufficient time to stabilize, but any later might allow faculty to drift from the career path they established in graduate school), but such data are not readily available. Instead, we can use digital versions of the directories of faculty that are published for most disciplines by the appropriate professional association. Here, we use a computer file of the American Political Science Association's (APSA) member directory, which includes Ph.D. information and current (as of 2005) faculty status for all registered faculty.[9] This contains information on all faculty in college and university departments affiliated with the APSA, including nearly all the major graduate departments. In order to keep a suitably large size to avoid too much statistical noise, but still produce relatively recent results, we restrict our data to faculty who were awarded the Ph.D. between 1990 and 2004.[10] Also, only faculty holding tenure-track professorial positions are included; lecturers, adjunct professors, postdoctoral fellows, and those in other positions are excluded.

Using a directory such as this also has significant advantages: since the (substantial) resources needed to maintain the data source for this ranking are already being invested for other reasons, what could be a very difficult exercise in data collection for the purpose of rankings becomes trivial. In addition, since learned societies in most academic disciplines maintain similar databases, the method is easily extensible to any number of other fields. Moreover, updates are planned regularly through the normal course of most learned societies' activities, making periodic follow-up rankings feasible.

As stated earlier, this ranking (as with any ranking) can only claim to measure one dimension of a program's quality. As a measure of program quality for aspiring graduate students, and those interested in graduate programs as sites of education (not just research), it has much to recommend it. The most appealing virtue for many will be its objectivity—unlike reputational surveys, the rankings are based entirely on real-world results, not on fallible agents' perceptions of quality.

The rankings themselves are displayed in Table 1, along with the NRC quality ranking from 1993. (*US News and World Report* scores are not reproduced for copyright reasons.) The first measure is the weighted influence ranking described in the Per-Capita Placement Success section; the second is the unweighted score described in the Basic Formula section. Both are run on the data set of all institutions in the APSA directory that awarded at least one Ph.D. in political science during the period under study (1990–2004).[11] We scale our scores from 0 to 100 (all relative to the highest scoring program).

Eighty-six schools are listed: those ranked in the two systems listed (ours and the National Research Council's) that awarded at least 30 Ph.D.s over the 15-year period covered. The schools are sorted by their score on the weighted measure, as it indicates which programs attract and train the highest-quality students on average, independent of the size of the program.

Interpretation of Results and Comparison to Other Ranking Systems

While reputational surveys have their shortcomings, it is not unreasonable to assume that the faculty members surveyed have at least a generally accurate perception of the quality of programs in their field. Thus they can provide a reality check for any other ranking system; a system that diverges too much from reputational measures can be assumed to have somehow "gotten it wrong" in the judgment of the majority of practitioners in a field.

Our ranking correlates fairly strongly with the existing rankings systems. The natural logarithm of our unweighted influence measure explains about 76% of the variation in the NRC rankings ($R^2 = 0.759$), which is substantial considering that the NRC rankings are over 13 years old.[12] By means of comparison, the NRC ran a multiple regression analysis of objective correlates of their Program Quality score in English programs using a variety of factors, such as number of faculty with awards, total number of doctorate recipients with academic employment plans, number of citations of program faculty, and other counts, which produced an R-squared value of 0.811 (Ostriker and Kuh 2003, 150).

Table 2 shows the correlation of our measure with the two most widely used measures: the NRC score and the *US News* ranking from 2005. Interestingly, the NRC and *US News* correlate slightly less strongly with our weighted influence measure than with our unweighted measure, which is suggestive evidence in favor of the hypothesis that program size is a determinant of recognition in reputational surveys.

The correlation with the metric of reputation should not be surprising. The strongest factor playing into this correlation is that prospective graduate students rely on reputations in deciding where to attend graduate school; that is to say, this method partially acts as a revealed-preference metric somewhat along the lines of Avery et al. (2004) by looking at which institutions the most talented undergraduates choose to attend. This is probably not the only factor driving these correlations. Although it is unclear to what extent graduate schools influence the future success of their graduates, having attended a perceived high-quality graduate school certainly would help a newly minted Ph.D. on the job market. In any case, the faculty who lead job search programs and those who fill out surveys about institutional quality probably have similar opinions, so it is

Table 1
Departmental Rankings Based on the Placement of Graduates Who Received their Ph.D.s between 1990 and 2004

School	Weighted Influence	Unweighted Influence	NRC Q Score
Harvard	100 (1)	100 (1)	4.88 (1)
Stanford	67.3 (2)	50.3 (3)	4.5 (5)
Michigan	53.2 (3)	47.8 (4)	4.6 (3)
Rochester	42.1 (4)	14.4 (12)	4.01 (11)
Chicago	39.7 (5)	46.7 (5)	4.41 (6)
California, Berkeley	38.3 (6)	59.4 (2)	4.66 (2)
Duke	35.7 (7)	20.1 (8)	3.94 (14)
UCLA	26.2 (8)	23.8 (6)	4.25 (8)
Northwestern	25.6 (9)	9.6 (15)	3.35 (22)
California, San Diego	24.8 (10)	12.3 (14)	4.13 (9)
MIT	24.6 (11)	22.2 (7)	3.96 (12)
Yale	24.4 (12)	19.1 (9)	4.6 (3)
Princeton	24.1 (13)	18.9 (10)	4.39 (7)
Cornell	21.6 (14)	12.6 (13)	3.85 (15)
Columbia	18.1 (15)	16.1 (11)	3.84 (16)
Washington, St. Louis	15.1 (16)	5.4 (17)	3.29 (24)
Michigan State	10.9 (17)	4.7 (19)	3.24 (26)
Ohio State	10.7 (18)	8.5 (16)	3.69 (17)
Emory	10.2 (19)	2.3 (31)	2.88 (36)
North Carolina, Chapel Hill	9.9 (20)	5.2 (18)	3.54 (18)
Penn	8.9 (21)	2.4 (30)	2.68 (42)
Florida State	8.6 (22)	3.1 (23)	2.82 (38)
Johns Hopkins	8.3 (23)	3.6 (21)	3.37 (21)
Brandeis	7.7 (24)	3.4 (22)	2.41 (53)
Washington	7.3 (25)	3.1 (23)	3.34 (23)
Oregon	6.9 (26)	1.9 (34)	2.21 (66)
Rutgers	6.6 (27)	4.7 (19)	3.24 (26)
SUNY, Stony Brook	6.4 (28)	2.8 (27)	2.92 (34)
Illinois, Urbana-Champaign	5.8 (29)	2.6 (29)	3.2 (30)
Boston College	5.2 (30)	1 (52)	2 (69)
Iowa	5.2 (30)	2.2 (32)	3.25 (25)
Minnesota	5.1 (32)	3 (25)	3.95 (13)
Indiana	5.1 (32)	2.7 (28)	3.45 (20)
Houston	4.8 (34)	2.2 (32)	2.96 (33)
Wisconsin-Madison	4.2 (35)	3 (25)	4.09 (10)
Southern California	3.7 (36)	1.5 (39)	2.33 (59)
Virginia	3.7 (36)	1.9 (34)	3.24 (26)
Pittsburgh	3.6 (38)	1.5 (39)	3.15 (31)
California, Irvine	3.6 (38)	1.1 (48)	3.14 (32)
Arizona	3.5 (40)	1.1 (48)	2.89 (35)
Rice	3.3 (41)	1.2 (45)	2.43 (51)
Texas, Austin	3.2 (42)	1.8 (36)	3.49 (19)
South Carolina	2.8 (43)	0.9 (54)	2.39 (54)
Colorado, Boulder	2.7 (44)	1.2 (45)	2.78 (39)
Syracuse	2.7 (44)	1.4 (41)	2.77 (40)
American	2.6 (46)	1.8 (36)	2.37 (56)
Boston U.	2.5 (47)	1.3 (42)	1.69 (74)
Vanderbilt	2.5 (47)	0.9 (54)	2.32 (62)
Grad. Center (CUNY)	2.5 (47)	1.3 (42)	2.57 (47)
California, Davis	2.4 (50)	1.3 (42)	2.61 (46)
Penn State	2.3 (51)	0.9 (54)	2.25 (65)
Louisiana State	2.2 (52)	0.9 (54)	2.02 (68)
Maryland, College Park	2.1 (53)	1.7 (38)	3.23 (29)
Connecticut	1.9 (54)	1.1 (48)	2.31 (63)
Georgia	1.9 (54)	1.1 (48)	2.66 (44)
Oklahoma	1.8 (56)	1.2 (45)	1.94 (70)
Arizona State	1.8 (56)	0.9 (54)	2.67 (43)
Florida	1.8 (56)	0.9 (54)	2.48 (50)
Kentucky	1.7 (59)	0.9 (54)	2.42 (52)

(continued)

Table 1 (Continued)

School	Weighted Influence	Unweighted Influence	NRC Q Score
SUNY, Binghamton	1.7 (59)	0.9 (54)	2.27 (64)
Clark Atlanta University	1.6 (61)	1 (52)	0.6 (86)
Tennessee, Knoxville	1.5 (62)	0.8 (64)	1.36 (81)
University of New Orleans	1.4 (63)	0.8 (64)	1.45 (79)
Notre Dame	1.4 (63)	0.9 (54)	2.66 (44)
Kansas	1.4 (63)	0.8 (64)	2.33 (59)
George Washington	1.4 (63)	0.8 (64)	2.57 (47)
SUNY, Buffalo	1.3 (67)	0.7 (71)	2.06 (67)
California, Santa Barbara	1.2 (68)	0.8 (64)	2.74 (41)
Washington State	1.2 (68)	0.8 (64)	1.39 (80)
Georgetown	1.2 (68)	0.8 (64)	2.85 (37)
Purdue	1.2 (68)	0.7 (71)	2.38 (55)
Claremont Grad.	1.1 (72)	0.9 (54)	1.8 (71)
North Texas	1.1 (72)	0.6 (74)	1.64 (76)
Hawaii	1 (74)	0.7 (71)	2.49 (49)
Massachusetts	0.9 (75)	0.6 (74)	2.37 (56)
Fordham	0.9 (75)	0.6 (74)	1.12 (84)
Missouri, Columbia	0.9 (75)	0.6 (74)	1.79 (72)
Howard	0.9 (75)	0.6 (74)	1.62 (77)
California, Riverside	0.8 (79)	0.6 (74)	2.36 (58)
Catholic	0.8 (79)	0.6 (74)	0.95 (85)
Cincinnati	0.8 (79)	0.6 (74)	1.65 (75)
Nebraska	0.8 (79)	0.6 (74)	2.33 (59)
Northern Arizona	0.8 (79)	0.6 (74)	1.17 (83)
Northern Illinois	0.8 (79)	0.6 (74)	1.77 (73)
Temple	0.8 (79)	0.6 (74)	1.54 (78)
Texas Tech	0.8 (79)	0.6 (74)	1.2 (82)

Note: State names without any other modifier refer to the flagship campus of the state university.

Table 2
Correlation Coefficients between Rankings[13]

Ranking	Weighted Influence	Unweighted Influence
NRC Q	0.860	0.912
US News	0.897	0.930

unsurprising that the same schools would be favored in hiring processes as in reputational surveys.

These overall high correlations show that using placement as a measure produces an ordering that largely conforms to the generally accepted definition of "quality" more closely than any other data-derived measure (the highest correlation with NRC Q score for any of the additional data points they provide in the most recent rankings is with faculty size, at 0.737; the next largest is 0.622, for citation count per faculty member). We consider it one of the strengths of our system that it generally corresponds to this accepted idea; but the differences between our system and reputational rankings are instructive as well. By focusing on only one dimension of quality (to the exclusion of other factors like faculty publication records) we create, as discussed above, rankings more closely attuned to the needs of prospective students. An additional important benefit is provided by the ability to discriminate, as reputational rankings cannot, between per-capita and unweighted measures. The weighted influence ranking in particular allows the recognition of high-quality but small programs that may be systematically disadvantaged in reputational surveys; schools like Rochester, Emory, and Duke score much higher on our weighted influence measure than they do on either our unweighted measure or on the NRC quality ranking.

No ranking system can adequately define a concept as multifaceted as the "quality" of graduate programs in political science, much less measure it precisely. Our method instead focuses on one crucially important aspect of graduate programs—the success of their graduates in the academic job market—and employs an "influence" metric that captures both the overall percentage of new Ph.D.s who hold tenure-track positions in political science as well as the influence of the institutions at which those jobs are held. A program's placement record reflects both the quality of the students it is able to attract as well as the training they receive, both of which should be of enormous interest to prospective graduate students as well as to the departments themselves. And as a tool for assessment it should aid both administrators and students in evaluating the likelihood of graduates obtaining a desirable job inside of academia, and, for many, the importance of making students aware of other options.

Critics of rankings often argue that a university may respond by taking steps counter to its mission in order to raise its position in the rankings. Unlike existing rankings, the only way departments could manipulate their performance on our metric would be to improve the placement success of their graduates, either by recruiting stronger students or better preparing them for the academic job market. The primary beneficiaries of such "manipulation" would be the students themselves.

Notes

*The authors extend their thanks to William Bowen, Derek Bruff, Jonathan Cole, Philip Katz, Gary King, Robert Townsend, Harriet Zuckerman, and two anonymous *PS* reviewers for their valuable comments on and criticisms of earlier drafts of this paper.

1. The method was first used to rank graduate programs in many disciplines by Allan Cartter (1966). The method was also used in subsequent studies by the American Council on Education, where Cartter was vice president (Roose and Anderson 1970) and then by the National Research Council (most recently, NRC 1995). In addition, quadrennial rankings of Ph.D. programs by the newsmagazine *US News and World Report* (most recently, *US News and World Report* 2006) rely entirely on reputational surveys.

2. It should be noted that not all of these complaints are entirely justified: while raters do tend to favor their own institutions of employment, and, even more heavily, the institutions at which they received their degrees (Cole and Lipton 1977, 666), a study commissioned by the American Association of Law Schools found no evidence of deliberate sabotage in the *US News* law school rankings (Klein and Hamilton 1998).

3. Masuoka et al. (2007) present rankings of political science departments based on the raw number of Ph.D.s placed at doctoral institutions in the U.S. They also find that a relatively small number of departments account for a disproportionately large number of placements. This result is consistent with the distribution of our ranking metric (only a few departments receive high scores, while the majority receive very low scores).

4. Research I universities are the largest and most prominent of the four doctorate-granting institutions in the Carnegie Classification. The doctoral programs ranked here are largely, but not exclusively, at Research I or Research II universities; some are at Doctoral I or II universities.

5. However, these data are somewhat dated—survey respondents received their Ph.D.s in the mid-1980s.

6. Google's algorithm, in turn, drew from a large amount of literature on academic citation analysis. See http://dbpubs.stanford.edu:8090/pub/1999-66 for the original Google paper. Our description of Google's algorithm also benefited from the Wikipedia entry on PageRank, at http://en.wikipedia.org/wiki/PageRank, accessed July 1, 2005.

7. This addition is used in Google's PageRank algorithm as well. There it represents the model of a random web surfer who follows links but occasionally makes a random leap to avoid getting "stuck" on a page with no links; here it serves much the same purpose, giving a base probability to every school that it will be randomly switched to.

8. The process can be conceptualized essentially the same way as described in the previous section, except that instead of choosing professors randomly at each school, it is weighted toward choosing professors from smaller schools.

9. We gratefully acknowledge Michael Brintnall and the American Political Science Association for sharing these data. The dataset is archived with the APSA.

10. Since the median school in these rankings awards about five degrees a year, the inclusion or exclusion of a single graduate can make an appreciable difference in the rank of schools in the lower half of the rankings. With a 15-year horizon, however, the general correlation of year-to-year scores is quite high ($r = 0.999$ for one year, $r = 0.985$ over four years).

11. Data on the number of Ph.D.s awarded, which are also used in our weighted influence ranking, are from the Integrated Postsecondary Education Data System (IPEDS).

12. There is an approximately logarithmic relationship between our scores and scores from the existing rankings, so all correlations are linear ones from the reputational surveys to the natural logarithm of our raw scores.

13. These coefficients are calculated from the set of 53 schools that are ranked by NRC, *US News*, and our method. *US News* only makes data available for schools scoring at least 2.5 (on a 5-point scale) in their ranking. For the set of 86 schools ranked by NRC and our method, the correlation coefficients between the NRC score and our scores are 0.853 for weighted and 0.871 for unweighted.

References

Avery, Christopher, Mark Glickman, Caroline Hoxby, and Andrew Metrick. 2004. "A Revealed Preference Ranking of U.S. Colleges and Universities." Available at http://post.economics.harvard.edu/faculty/hoxby/papers/revealedprefranking.pdf.

Cartter, Allan M. 1966. *An Assessment of Quality in Graduate Education*. Washington, D.C.: American Council on Education.

Cole, Jonathan R., and James A. Lipton. 1977. "The Reputations of American Medical Schools." *Social Forces* 55 (3): 662–84.

Hix, Simon. 2004. "A Global Ranking of Political Science Departments." *Political Studies Review* 2: 293–313.

Hoffer, T. B., V. Welch, Jr., K. Williams, M. Hess, K. Webber, B. Lisek, D. Loew, and I. Guzman-Barron. 2005. *Doctorate Recipients from United States Universities: Summary Report 2004*. Chicago: National Opinion Research Center. (The report gives the results of data collected in the Survey of Earned Doctorates, conducted for six federal agencies, NSF, NIH, USED, NEH, USDA, and NASA by NORC.)

Ingram, Linda, and Prudence Brown. 1997. *Humanities Doctorates in the United States: 1995 Profile*. Washington, D.C.: Office of Scientific and Engineering Personnel, National Academy Press.

Klein, Stephen P., and Laura Hamilton. 1998. "The Validity of the U.S. News and World Report Ranking of ABA Law Schools." American Association of Law Schools online publication, available at www.aals.org/reports/validity.html. Accessed August 15, 2005.

Laband, David N. 1986. "A Ranking of the Top U.S. Economics Departments by Research Productivity of Graduates." *Journal of Economic Education* 17 (1): 70–6.

Masuoka, Natalie, Bernard Grofman, and Scott L. Feld. 2007. "The Production and Placement of Political Science Ph.D.s, 1902–2000." *PS: Political Science and Politics* 40 (April): 361–6.

McCormick, James M., and E. Lee Bernick. 1982. "Graduate Training and Productivity: A Look at Who Publishes." *Journal of Politics* 44 (1): 212–27.

National Research Council. 1995. *Research-Doctorate Programs in the United States: Continuity and Change*. Washington, D.C.: National Research Council.

Nerad, Maresi, and Joseph Cerny. 2003. "Career Outcomes of Political Science Ph.D. Recipients: Results from the Ph.D.s Ten Years Later Study." Seattle: Center for Research and Innovation, University of Washington, available at http://depts.washington.edu/coe/cirge/pdfs%20for%20web/Poli%20Sci%20Report6.pdf.

Ostriker, Jeremiah P., and Charlotte V. Kuh, eds. 2003. *Assessing Research—Doctorate Programs: A Methodology Study*. Washington, D.C.: National Academies Press.

Page, Lawrence, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1998. "The PageRank Citation Ranking: Bringing Order to the Web." Available at http://dbpubs.stanford.edu:8090/pub/1999-66.

Roose, Kenneth D., and Charles J. Anderson. 1970. *A Rating of Graduate Programs*. Washington, D.C.: American Council on Education.

United States Department of Education, National Center for Education Statistics. "Integrated Postsecondary Education Data System (IPEDS), 1989–1990 through 2003–2004 [computer files]." Available at http://nces.ed.gov/ipeds/.

U.S. News and World Report, eds. 2006. "America's Best Graduate Schools 2006." Washington, D.C.: *U.S. News and World Report*. Additional data from premium online edition, www.usnews.com/usnews/edu/grad/rankings/rankindex.php.