ASSESSING DOCTORAL PROGRAM QUALITY: AN OVERVIEW OF THE 2009 NATIONAL RESEARCH COUNCIL REPORT

Background

The concern for assessing the quality of doctoral education in the United States extends back to the 1920's when Raymond Hughes, the then president of the University of Miami at Ohio, undertook a survey of twenty doctoral programs in 1925. While ground-breaking at the time for bringing attention to the quality of doctoral education in the United States, Hughes' work was criticized on methodological grounds.[1]

The assessment of doctoral program quality became a mainstay of educational research during the middle part of the twentieth century. Studies during this time were of three general types: reputational studies; objective indicator studies; and quantitative correlate studies. While reputational studies were the most popular, they also received the most criticism. Perhaps the most recognized and accepted of these studies during this period was the first National Research Council study, *An Assessment of Research-Doctorate Programs in the United States* (Jones, Lindzey, and Coggeshall, 1982).[2]

In an effort to update and expand the work of the first NRC study, and to improve upon the various deficiencies of the previous studies of doctoral program quality, the NRC undertook a second study of doctoral programs in the early 1990s. The report, *Research-Doctorate Programs in the United States: Continuity and Change* (Goldberger, Maher, and Flattau, 1995), represented at the time the most comprehensive response to the various critiques leveled against evaluations of doctoral program quality in the United States.

The 1995 study received a great amount of attention and ignited renewed interest in the assessment of doctoral program quality.[3] While the second NRC study was much

---

[1] Hughes asked well-respected scholars to review programs in their respected fields. Criticisms of his work focused on the small size of his panel of experts and the geographical imbalance among the scholars who rated the programs.

[2] The reliance of reputational studies on the subjective evaluation of doctoral programs by faculty and/or academic administrators (e.g., departmental heads and deans) to arrive at a rating for programs within an academic field was seen more as an assessment of a program's faculty reputation rather than an indicator of a program's overall quality. Reviews of those studies that occurred during this period can be found in Conrad and Blackburn (1985) and Tan (1986). Many criticisms, both methodological and conceptual, have been leveled against all three types of studies. In addition to the criticism that many of these studies, and especially reputational studies, tend to confuse the reputation of a doctoral program for quality, other criticisms that have been directed at these studies include the existence of: rater bias; outdated data; unreliable and incorrect data; unidimensionality (i.e., the lack of including more than one indicator of quality); and an atheoretical focus (this latter criticism is often cited against correlate studies). Although these critiques tell us that we should be careful when interpreting the results of these studies, they also have contributed to an ongoing refinement of studies in the assessment of doctoral education.

[3] Articles in Change magazine by Webster and Skinner (1996) and Ehrenberg and Hurst (1996), in higher education news media such as the Chronicle of Higher Education (Magner, 1995; Graham and Diamond,

improved over the earlier 1982 study, its reliance on reputational evaluations of doctoral programs by faculty raters became the dominant focus for those both inside and outside of academia. Rather than becoming a tool for internal assessment and improvement, the 1995 NRC study quickly took on the feel of a beauty pageant, with the winning institutions being those with the highest number of programs rated highly by faculty raters in some top percentage range. Nevertheless, the 1995 NRC study and its many reviews set the stage for the third and most comprehensive incarnation of the NRC's evaluation of doctoral programs, *Assessment of Doctorate Programs* (2009).[4]

Data and Methods[5]

The most recent NRC evaluation of doctoral programs expands and improves on the 1995 study in a variety of ways. The first obvious difference with the 1995 study is the sheer size of the undertaking. The 1995 report included over 3,500 programs in 41 fields of study at 274 universities compared to the most recent NRC study that included more than 5,000 doctoral programs across 61 fields of study in 222 institutions.[6] Taxonomic variations among the disciplines represent another noted difference between the second and third NRC studies.[7] The 2009 report was expanded to include Nursing and more life science fields, and a few other changes and additions in the Arts and Humanities, Social Science, and Physical Sciences and Engineering were also made.[8] But perhaps the most important difference between the 1995 and 2009 reports resides in the data that were included and how the data were used to develop the weights generating the rankings presented in the study.

<u>Data Sources</u>

Data used in the 2009 NRC report came from primary data collection instruments and secondary data sources. The study used five surveys administered to different populations to obtain data needed to conduct the latest NRC evaluation of doctoral programs. These surveys included:

- An institutional questionnaire that asked for a list of doctoral programs at each institution and questions that asked about institution-specific practices.

1999), and independent analysis of the data collected in the 1995 study (Ehrenberg and Hurst, 1998; Junn and Books, 2000; Lorden and Martin, n.d.; Maher, 1996; Toutkoushian, Dundar, Becker, 1998) all illustrated the tremendous interest that emerged from the publication of the second NRC study.

[4] The 2003 report, *Assessing Research-Doctorate Programs: A Methodology Study*, provided the conceptual and methodological framework for the 2009 NRC report.

[5] The following discussion is a distillation of much of what is presented in *A Guide to the Methodology of the National Research Council Assessment of the Doctorate Programs (2009)*.

[6] Although the number of doctoral granting institutions has increased by over 10% between the time that the second and third NRC studies were conducted, the decrease in the number of participating institutions in the 2009 report is likely due to the requirement that universities had to pay to submit data and be listed in the most recent study. Institutions were assessed fees ranging from $5,000 to $20,000 depending on the average number of study-relevant Ph.D.s granted (based on 2001 to 2003 data).

[7] The criteria for a field or discipline to be included in the study were that at least 500 Ph.D.s were produced over a five year period prior to 2004-2005 in at least 25 universities.

[8] See the taxonomy chart showing these differences between the 1995 and 2009 reports.

- A questionnaire for each program at an institution that was to be included in the study – each program questionnaire asked questions about students, faculty, and program characteristics. Each program was asked to provide a list of their doctoral faculty.[9] Five programs – one in each broad discipline area (English, chemical engineering, economics, physics, and neuroscience) – were also asked to provide a list of their advanced doctoral students.
- A faculty questionnaire that asked individual faculty members about their educational, work, research, and publication history. An important part of this questionnaire was a section that asked faculty to identify those factors which they thought were critical to the quality of doctoral education in their field. Faculty were also asked if they would be willing to rate other programs within their field.
- A student questionnaire sent to advanced doctoral students in the fields noted above. Questions focused on student educational background, experiences they had while in the program, including research activities, and their post-graduation plans.[10]
- A rating questionnaire sent to a stratified sample of faculty who agreed to be raters of doctoral programs in their field.

Data were also collected from secondary sources. Data on publications and citations were collected for faculty in all fields except the humanities.[11] Data on faculty honors were obtained from 224 scholarly societies for all fields, with a distinction made between highly prestigious awards and all other awards. Another key source for data was the Survey of Earned Doctorates (SED) that all Ph.D. candidates complete when filing for their degree.

The collected data provided the basis for constructing 20 “dimensional” measures which were divided into three categories: research activity, student support and outcomes, and diversity of the academic environment.[12] These three broad categories were seen as essential aspects of doctoral program quality and as such were considered critical factors for understanding and assessing the quality of doctoral programs in the United States today.[13]

---

[9] Respondents were asked to divide their faculty into core faculty (faculty actively supervising doctoral candidates or serving on admissions or curriculum committees), new faculty, and associated faculty (faculty who work at the institution and are supervising dissertations but are not considered core faculty).

[10] The present NRC study sought to include student input in the belief that direct assessment of student experiences provides critical –and unique - information about program quality. The data collected from students were not used in the rating of doctoral programs.

[11] These data were collected through the Institute of Scientific Information (ISI) and matched to lists of faculty for fields in the sciences and social sciences. The citation count was based on the years 2000-2006 and relates to publications between 1981 and 2006. Data for humanities faculty came from counts of books and publications going back to 1996; these counts were based on evaluations of each faculty member’s curriculum vitae.

[12] See the accompanying table for a listing of variables used in the NRC study.

[13] Obviously, all these factors, or dimensions, have a numerical quality, and therefore, lend themselves to quantitative manipulation. Other considerations (e.g., the creative use of resources, the blending of disciplines, or the existence of a scholarly community) most assuredly affect doctoral program quality, but are unfortunately not as easily measurable and therefore, were not included in the NRC study.

Methodology

At the heart of the 2009 NRC report on doctoral program quality is the effort to construct weights that allow for the projection of a range of rankings for each program included in the study. These weights and range of rankings are constructed for the three categories of variable dimensions: research activity, student support and outcomes, and program diversity. An overall set of weights and range of rankings are also constructed for each evaluated program.

An important finding about each of the three dimensional categories of weights and ratings is that they produce program rankings that are different from the overall rankings of programs. The only exception to this is the faculty research activity dimension, which has a ranking that is similar to the ranking produced by all twenty variables and reflects the widespread consensus about the importance of faculty productivity for program quality. The differential impact of the dimensional categories on program weighting and ranking supports the notion that doctoral program quality is not unidimensional and quality may mean different things to different individuals - any assessment of a program must take into account the wide range of factors that define program quality. For example, students selecting doctoral programs may find the practicalities of program involvement (such as a relatively short time to degree, high completion rates, and stable funding) more desirable than 'prestigious' programs, which are often defined by high faculty research activities.

As the report's methodology guide points out,[14] the present NRC study sought to accomplish two fundamental goals in response to the perceived shortcomings of the 1995 NRC report:

- to include as much information as possible about the programs being rated so that each can be examined and understood in a broader context than previously; and
- to mitigate as much uncertainty about the methodology of rating doctoral programs as possible.

The former is seen as being addressed with the construction and dissemination of dimensional weights and rankings while the second point is addressed through the statistical practice of resampling. What follows is an overview of how the weights used in the NRC study are constructed and how resampling is used to reduce the uncertainty surrounding their construction and application in the rankings of doctoral programs.

Estimation and Uncertainty in the Assessment of Doctoral Program Quality

A major criticism of the 1995 NRC report centered on how its reputational surveys were used to construct ratings and rankings. Rather than seeing ratings and rankings of programs as estimates that could vary depending on the programs evaluated and the raters

[14] See, supra., note 5.

selected, the ordinal presentation of the summative rankings in the 1995 NRC report were instead taken to be precise estimates of the position of doctoral programs in each field.[15] A fundamental task of the NRC researchers conducting the 2009 study was to incorporate this variance, or uncertainty, into the estimation of weights and rankings of programs in their methodology and data presentation. They did this by utilizing a resampling technique known as Random Halves, which is seen as accounting for uncertainty by empirically building a distribution of possible rankings for each program.[16]

Step 1

The program weights used in the NRC study were the combination of two different sources of data. From the faculty survey, which asked faculty to indicate the relative importance of each of the 20 factors of program quality in their designated field, a 'direct' weight for each factor of program quality was constructed within each of the 61 disciplines.[17] This approach, noted as a "bottom-up" method in building ratings, gives importance to what faculty consider affecting program quality without any reference to specific institutional programs.

However, the calculation of these weights are viewed to have a level of built-in uncertainty stemming from the likelihood that a different set of faculty would probably result in a different set of weights. It is very unlikely that all faculty within a discipline would agree unanimously on the relative importance of each factor considered. Consequently, the random halves resampling procedure was employed to construct a distribution of plausible weights that could have been constructed if different faculty were included. Researchers re-sampled the sample of faculty who answered questions on

[15] Confidence intervals were presented in the second NRC report, underscoring the observation that these rankings were indeed estimates and that the ratings and rankings did have a degree of variance. However, this basic fact was simply ignored in how the rankings in the 1995 NRC report were used.

[16] Resampling methods that include random halves attempt to establish the range of possible values of what is being observed given different possibilities in the data being observed. It does not try to infer if what is being observed actually exists in some real or hypothetical population. (Hence, the resampling approach is essentially a nonparametric statistical method for arriving at answers to probabilistic questions.) Random halves is a modified version of the jackknife method of resampling. Another resampling technique considered by the NRC researchers was the bootstrap technique. Random halves was selected over bootstrapping because it only required the resampling of half of the sample and uses sampling without replacement. Bootstrapping is a much more resource intensive technique that requires the resampling of the entire sample with replacement. For more information on resampling techniques see Good (2005, 2006), Mooney and Duval (1993), and Simon (1977).

[17] Faculty within a discipline were asked to select the top four factors of importance within each of the three dimensional categories. Each factor selected was assigned a value of 1. Faculty were then asked to designate the two most important from the four factors previously selected. Again a value of 1 was assigned to the selected factors. Faculty were also asked to assign a weight to each dimensional category, ranging from 0 to 100 with the total weight not exceeding 100. A value was then calculated for each of the 20 factors by adding the values of the first two steps and multiplying by the weight given to the dimensional category each factor belongs to. The total value across all twenty factors adds to 600. The relative weight of each factor was then transformed to a scale between 0 and 1 by dividing the product of each factor (the addition of the 1s multiplied by the weight assigned between 0 and 100) by the total value (600). The average weight for all faculty within a program represents the direct measure, or weight, for each of the twenty factors.

the faculty survey about the relative importance of the 20 factors using the random halves procedure. The procedure was replicated 500 times to obtain an empirically-derived distribution of direct weight vectors.[18] Thus, by resampling, a distribution of weights was generated and used to represent the range of plausible weights produced from surveying faculty.

Step 2

The second source for program weights involved the application of regression analysis to a sample of programs rated by a random group of faculty raters within each field. This regression-based method, sometimes referred to as a "top-down" approach, differs from the direct approach noted above by starting with ratings of actual programs and then applying statistical techniques to infer weights to the program variables from the evaluations given by a sample of raters.

Because the number of programs within most fields was quite high, often approaching if not exceeding 100, it would have been too exhaustive to have raters evaluate each program. In fields with a large number of programs, 50 programs were sampled at random from a stratified classification of programs based on program size and geographical location. Fields with a smaller number of programs resulted in a stratified sample of 30 programs. Stratified sampling was applied to the population of faculty agreeing to serve as raters of programs in their field based on size of their program, geographical location of their program, and their faculty rank. Each rater was asked to evaluate a set of 15 randomly selected programs using a 6-point scale from 1 ("not adequate for doctoral education") to 6 ("distinguished"), resulting in approximately 50 assessments for each program in the field.[19] An average for each rated program was then calculated and used as the "dependent" variable in a regression model that related the average rating for each program in a discipline to the data on each of the program factors.[20] The resulting coefficients were then transformed to the same scale as the direct weights (between 0 and 1).[21]

---

[18] Each simulation resulted in a set of weights across the 20 program factors.
[19] Each program, on average, was rated by approximately 6 out of 10 raters who were given data about the program. Non-respondents were replaced by other raters from the same stratum until almost every program was rated by approximately 50 raters.
[20] Because of the low number of cases (50 programs) in relation to the number of "independent" variables (20) and the high degree of multicollinearity among the program variables, principle component variables were created from the original standardized program variables. The average rating of each program was then regressed on the principal component variables using a backward, step-wise selection method to eliminate those principal component variables that had a "non-significant" contribution to the variance of the program ratings. This procedure thus reduced the number of variables included in the final regression model, thereby enhancing the robustness of the regression model. When the best regression model was obtained, the calculated regression coefficients of remaining principal component variables were transformed back to the scale of the original program variables. Because the derived principal component variables are linear combinations of the original twenty program variables that are uncorrelated and can yield the same predictions as the original variables, the transformation of the regression coefficient results in each of the 20 original variables retaining a non-zero weight.
[21] The absolute value of each regression coefficient was added together and then the regression coefficient for each program variable, or factor, was divided by that total.

As with the procedure used for estimating the vector of direct weights above, random halves resampling was applied to the sample of programs and raters used in the regression-based weights. The administration of random halves resampling is seen as reducing the uncertainty that comes from sampling both the faculty raters and the programs rated.[22] The procedure sampled a random half of all raters for programs in a discipline and computed the average rating of each program from that half-sample. The regression of the average rating on the twenty program variables across the 50 sampled programs was then replicated 500 times to obtain an independent distribution of 500 regression weight vectors. This distribution represents the bounds within which fall the weights produced by various combinations of faculty raters and programs.

Step 3

The 500 replications of direct weights and the 500 replications of regression-based weights were then combined into 500 replications of combined weights that then became the basis for estimating the distribution of scores for each program within a discipline.[23] This distribution of combined weights was then used to estimate the rating values for each program in a field as noted in step 4.[24]

Step 4

Just as there was uncertainty in the calculation and combination of the two sets of weights, there is also uncertainty in the values of the program variables themselves. The 20 program variables used to calculate the ratings can vary or have an error associated with their values because of year to year fluctuations.[25] To account for this uncertainty an error factor was applied to each program variable.[26] Rather than using the fixed values of the program variables, each value of an assessed program was perturbed 500 times and standardized for each of the 20 program variables.[27] These 500 perturbed standardized values for each program across all 20 program variables were then

[22] Such uncertainty also existed in the 1995 NRC study as well, but because the rankings were presented as specific scores obtained by averaging questionnaire responses they were ignored as estimates of rankings, even though confidence intervals were included in the 1995 report.

[23] The procedure for combining these two weight vectors involved the use of a "policy weight," which determines the degree to which both the direct and regression-based weights influence their combination, and the minimization of error in the combination of the weights.

[24] If the resulting distributions of combined weights for a program variable contained a zero that fell inside the middle 95 percent of its distribution then that program variable was considered insignificant and removed from further computations. Such elimination of program variables resulted in the re-calculation of all previous steps, with the eliminated program variables being ignored in the new calculations of direct and regression-based weights.

[25] Certain program variables are based on data collected over a number of years, where averages over the time period were calculated and used. One source of variability thus becomes the time frame used: different time periods would likely result in different values.

[26] This factor was determined either by the standard error of the variable (SD/square root of N) divided by the value of the program variable for variables that were based on data collected over a period of time or by the assignment of a relative error of 0, .1, or .2 as determined by the NRC researchers.

[27] Perturbation of each program factor's value was accomplished through drawing randomly from the Gaussian distribution.

combined with the 500 replications of weights to obtain 500 ratings for each program. These ratings were ranked to obtain the inter-quartile range of the rankings for each program.

Presentation of Results

Results from the most recent NRC study of research doctorate programs contain data that facilitate benchmarking and analysis of individual programs across institutions within a discipline. Programs are ranked and compared according to their performance on three variable dimensions: research activity, student support and outcomes, and diversity of the academic environment. These three dimensions also form the basis for the overall assessment and ranking of individual programs within each field. However, in an attempt to promote the understanding that such rankings are estimates of the ordering of doctoral program quality within a discipline, and do not have the precision that is often attributed to singular rankings of programs, the 2009 NRC report provides a range of plausible rankings for each individual program. Each program in the ranking of a particular field is given a low and high ranking. Such a range suggests that programs can overlap and that a program can possibly take on the value of any number of rankings. Thus programs are not only ranked on different characteristics (research activity, student support/outcomes, and diversity) in addition to an overall listing, but each program is presented as likely falling within some interval of programs rather than a fixed place among all programs.

More specifically, 500 simulated rankings for each program were produced using the combined weights and perturbed values for the twenty variables identified as affecting doctoral program quality. These 500 rankings were then ordered from highest to lowest and the interquartile range of this ordering is presented for each program. Reporting the data in this way precludes the possibility of comparing the 2009 assessment results with past NRC studies when only a single estimate was presented of a program’s ranking. The presentation of quartile rankings makes it apparent that there are many plausible rankings for a program, and that the ‘real’ ranking for a program is likely to fall within the range presented.

One way to benchmark against other programs listed in the 2009 NRC report is to compare the quartile rankings of a program with the quartile rankings of comparable programs. The number of programs that fall above (or below) a program’s quartile rank can be counted so that the placement of that program in comparison to other programs can be estimated.

Another distinctive feature of the reporting of results found in the 2009 NRC report is the amount of data presented and available for internal assessment. Presenting rankings based on different dimensional qualities indicates in which areas programs are performing well or poorly. The NRC study found that the various dimensions had a differential impact on the overall ranking of programs in many fields, with research activity tending to have the largest impact of the dimensions. Thus a program may do well on the overall ranking of programs in a field, but may be discovered to not do as well on certain dimensions such as student support or diversity. Or it could be the other way around, where a program does well on these dimensions, but performs less well on the research activity dimension and thus does not have a particularly high ranking overall.

Data are also provided on each of the twenty variables that were used in the NRC study to construct the different weights and subsequent rankings for each program in each field. The combination of a program's values on the twenty program variables and the range of weights assigned to these variables can be used to assess the individual strengths and weaknesses of specific programs.

SELECTED REFERENCES

Clark, M.J., R.T. Hartnett, and L.L. Baird, 1976. *Assessing the Dimensions of Quality in Doctoral Education: A Technical Report on a National Study in Three Fields.* Princeton, N.J.: Educational Testing Service

Conrad, C.F. and R.T. Blackburn, 1985. "Program Quality in Higher Education: A Review and Critique of Literature and Research." In *Higher Education: Handbook of Theory and Research* Vol. I, pp. 283-308. New York: Agathon Press.

Jones, L.V., G. Lindzey, and P.E. Coggeshall, 1982. *An Assessment of Research-Doctorate Programs in the United States*. Five Volumes. Washington, D.D.: National Academy Press.

Ehrenberg, R.G. and P.J. Hurst, 1996. "The 1995 NRC Ratings of Doctoral Programs," *Change* Vol. 28, No. 3.

Goldberger, M.L., B.A. Maher, and P.E. Flattau, 1995. *Research-Doctorate Programs in the United States: Continuity and Change.* Washington, D.C.: National Academy Press.

Good, P. 2005. *Introduction to Statistics Through Resampling Methods and Microsoft Office Excel*. New York: Wiley.

Good, P., 2006. *Resampling Methods*. 3rd Ed. Boston: Birkhauser.

Graham, A.G., R.W. Lyman, and M. Trow, 1995. *Accountability of Colleges and Universities: An Essay.* New York, N.Y.: Columbia University.

Graham, H.D., and N. Diamond, 1999. "Academic Departments and the Ratings Game," *The Chronicle of Higher Education,* June 18:B6.

Junn, J., and R. Brooks, 2000. "A Brief Analysis of the 1992-93 Reputational Data in the 1995 NRC Report," Report submitted to the NRC, September 14.

Lorden, J., and L. Martin, n.d. "Toward a Better Way to Rate Research Doctoral Programs. A position paper prepared for the NASULGC's Council on Research Policy and Graduate Education."

Magner, D.K., 1995. "Long-Awaited Survey Ranks 3,634 Doctoral Programs Nationwide," *The Chronicle of Higher Education,* September 22:A20.

Maher, B.A., 1996. *How to Read the 1995 National Research Council Report Research-Doctorate Programs in the United States: A Guide for Students and Others*. Washington, D.C.: National Academy Press.

Mooney, C. Z., and R.D. Duval, 1993. *Bootstrapping. A Nonparametric Approach to Statistical Inference*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-095. Newbury Park, CA: Sage.

Ostriker, J.P., P.W. Holland, C. V. Kuh, and J. A. Voytuk, editors, 2009. *A Guide to the Methodology of the National Research Council Assessment of the Doctorate Programs*. Committee to Assess Research- Doctorate Programs. National Research Council.

Simon, J. L. (1997). Resampling: The New Statistics. Retrieved from http://www.resample.com/content/text/index.shtml.

Tan, D.L., 1986. "The Assessment of Quality in Higher Education: A Critical Review of the Literature and Research," *Research in Higher Education* Vol. 24, No. 3.

Toutkoushian, R. K., H. Dundar, and W.E. Becker, 1998. "The National Research Council graduate program ratings: What are they measuring?," The Review of Higher Education, 21, 427-443.

Webster, D.S. and T. Skinner, 1996. "Rating Ph.D. Programs," *Change* Vol. 28, No. 3.

Comparison of 1995 and 2009 NRC Taxonomies

		Number of Rated Programs in NRC Studies	
1995	2009	1995	2000
LIFE SCIENCES			
Biochemistry and Molecular Biology		194	
	Biochemistry, Biophysics, and Structural Biology		159
Cell and Developmental Biology (incl. Microbiology)	Cell and Developmental Biology	179	122
Ecology, Evolution, and Behavior		129	
	Ecology and Evolutionary Biology		94
	Public Health		92
Molecular and General Genetics		103	
	Genetics and Genomics		65
	Immunology and Infectious Disease		78
	Biology/ Integrated Biology/ Integrated Biomedical Sciences		120
	Kinesiology		41
Cell and Developmental Biology (incl. Microbiology)	Microbiology		74
Neurosciences		102	
	Neuroscience and Neurobiology		94
	Nursing		52
Pharmacology		127	
	Pharmacology, Toxicology and Environmental Health		116
Physiology	Physiology	140	63
	Animal Sciences		60
	Entomology		28
	Food Science		31
	Forestry and Forest Sciences		33
	Nutrition		44
	Plant Sciences		

Comparison of 1995 and 2009 NRC Taxonomies

1995	2009	Number of Rated Programs in NRC Studies: 1995	Number of Rated Programs in NRC Studies: 2000
			116
	Emerging Fields:		
	Bioinformatics – Not Rated		17
	Biotechnology – Not Rated		4
	Systems Biology – Not Rated		2
PHYSICAL SCIENCES, MATHEMATICS AND ENGINEERING			
	Applied Mathematics		33
Astrophysics and Astronomy	Astrophysics and Astronomy	33	34
Chemistry	Chemistry	168	178
Computer Sciences	Computer Sciences	108	126
Geosciences	Earth Sciences	100	140
Mathematics	Mathematics	139	127
Oceanography		26	
	Oceanography, Atmospheric Sciences and Meteorology		50
Physics	Physics	147	160
Statistics and Biostatistics		65	
	Statistics and Probability		61
Aerospace Engineering	Aerospace Engineering	33	31
Biomedical Engineering	Biomedical Engineering and Bioengineering	38	74
Chemical Engineering	Chemical Engineering	93	106
Civil Engineering	Civil and Environmental Engineering	86	130
	Computer Engineering		20
Electrical Engineering	Electrical and Computer Engineering	126	136
	Engineering Science and Materials		12
Materials Science	Materials Science and Engineering	65	83

Comparison of 1995 and 2009 NRC Taxonomies

		Number of Rated Programs in NRC Studies	
1995	2009	1995	2000
Mechanical Engineering	Mechanical Engineering	110	127
Industrial Engineering		37	
	Operations Research, Systems Engineering and Industrial Engineering		72
	Emerging Fields:		
	Computational Engineering – Not Rated		4
	Information Science - Not Rated		19
	Nanoscience and Nanotechnology – Not Rated		8
	Nuclear Engineering – Not Rated		19
SOCIAL AND BEHAVIORAL SCIENCES			
	Agricultural and Resource Economics		28
Anthropology	Anthropology	69	82
	Communication		83
Economics	Economics	107	117
Geography	Geography	36	49
Linguistics	Linguistics	41	52
Political Science	Political Science	98	105
	Public Affairs, Public Policy and Public Administration		54
Psychology	Psychology	185	236
Sociology	Sociology	95	118
	Emerging Fields:		
	Criminology and Criminal Justice – Not Rated		14
	Science and Technology Studies – Not Rated		5
	Urban Studies and Planning – Not Rated		23
ARTS AND HUMANITIES			
	American Studies		22
Classics	Classics	29	

Comparison of 1995 and 2009 NRC Taxonomies

		Number of Rated Programs in NRC Studies	
1995	**2009**	**1995**	**2000**
			31
Comparative Literature	Comparative Literature	44	46
English Language and Literature	English Language and Literature	127	119
French and Francophone Language and Literature	French and Francophone Language and Literature	45	43
German Language and Literature	German Language and Literature	32	29
	Language, Societies, and Cultures – Not Rated		94
History	History	111	137
Art History		38	
	History of Art, Architecture and Archaeology		58
Music	Music	65	63
Philosophy	Philosophy	72	90
Religion	Religion	38	40
Spanish and Portuguese Language and Literature	Spanish and Portuguese Language and Literature	54	60
	Theatre and Performance Studies		27
	Emerging Fields:		
	Feminist, Gender, and Sexuality Studies – Not Rated		8
	Film Studies – Not Rated		7
	Race, Ethnicity and post-Colonial Studies – Not Rated		9
	Rhetoric and Composition – Not Rated		9

The Twenty Variables Used in the 2009 NRC Assessment of Research-Doctorate Programs

Research Activity Variables:

1. Publications per allocated faculty members, 2001-2006 (from ISI for non-humanities and books from faculty CVs in humanities)
2. Average citations per publication (from ISI – non-humanities)
3. Percent of core and new doctoral faculty respondents holding grants
4. Honors and awards per allocated faculty member
5. Interdisciplinarity as measured by the percent of associated faculty

Student funding and Outcomes Variables:

6. Average GRE score for 2004-06 (verbal for humanities, quantitative for all other disciplines)
7. Percent of students with full support in the 1st year
8. Percent of first year students with external funding
9. Average annual PhDs graduated 2002-2006
10. Average completions (8 years for humanities, 6 years for others)
11. Time to degree for full- and part-time students
12. % PhDs with definite plans for an academic position, 2001-05 (including post-docs), based on the Survey of Earned Doctorates
13. Student Work Space (1=100% of students w/workspace, -1 if<100% students w//workspace)
14. Health insurance (1=provides health insurance, -1=does not)
15. Student Activities (number offered from a list of 18)

Diversity Variables:

16. Percent of non-Asian minority core or new faculty
17. Percent of female core or new faculty
18. Percent of non-Asian minority students
19. Percent female students
20. Percent international students