A Revised Guide to the Methodology of the Data-Based Assessment of Research-Doctorate Programs in the United States (2010)

Jeremiah P. Ostriker, Paul W. Holland, Charlotte V. Kuh, and James A. Voytuk, editors; Committee to Assess Research-Doctorate Programs; National Research Council

ISBN: 0-309-15929-6, 65 pages, 8 1/2 x 11, (2010)

A Revised Guide to the Methodology of the Data-Based Assessment of Research-Doctorate Programs in the United States (2010)

Jeremiah P. Ostriker, Paul W. Holland, Charlotte V. Kuh, and James A. Voytuk, editors

Committee to Assess Research-Doctorate Programs

Board on Higher Education and Workforce
Policy and Global Affairs

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

PREPUBLICATION COPY—UNEDITED PROOFS

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, N.W. Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This project was supported by the Andrew W. Mellon Foundation, the Alfred P. Sloan Foundation, the U.S. Department of Energy (Grant DE-FG02-07ER35880), the National Institutes of Health (Grant N01-OD-4-2139, TO#170), the National Science Foundation (Grant OIA-0540823), the National Research Council, and contributions from 212 U.S. universities.

Additional copies of this report are available from The National Academies Press, 500 Fifth Street, N.W., Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, http://www.nap.edu.

Suggested citation: National Research Council. 2010. A Revised Guide to the Methodology of the Data-Based Assessment of Research- Doctorate Programs in the United States. Washington, DC: The National Academies Press.

Printed in the United States of America.

PREPUBLICATION COPY—UNEDITED PROOFS

Committee to Assess Research-Doctorate Programs

Jeremiah P. Ostriker, *Committee Chair,* Charles A. Young Professor of Astronomy and Provost Emeritus, Princeton University
Virginia S. Hinshaw, *Vice Chair,* Chancellor, University of Hawai'i at Mānoa
Elton D. Aberle, Dean Emeritus of the College of Agricultural and Life Sciences, University of Wisconsin–Madison
Norman Bradburn, Tiffany and Margaret Blake Distinguished Service Professor Emeritus, University of Chicago
John I. Brauman, J. G. Jackson–C. J. Wood Professor of Chemistry, Emeritus, Stanford University
Jonathan R. Cole, John Mitchell Mason Professor of the University, Columbia University (resigned June 2010)
Paul W. Holland, Frederic M. Lord Chair in Measurement and Statistics (retired), Educational Testing Service
Eric W. Kaler, Provost and Senior Vice President for Academic Affairs, Stony Brook University
Earl Lewis, Provost and Executive Vice President for Academic Affairs and Asa Griggs Candler Professor of History and African American Studies, Emory University
Joan F. Lorden, Provost and Vice Chancellor for Academic Affairs, University of North Carolina at Charlotte
Carol B. Lynch, Dean Emerita of the Graduate School, University of Colorado, Boulder
Robert M. Nerem, Parker H. Petit Distinguished Chair for Engineering in Medicine, and former director, Parker H. Petit Institute for Bioengineering and Bioscience, Georgia Institute of Technology
Suzanne Ortega, Provost and Executive Vice President for Academic Affairs, University of New Mexico
Robert J. Spinrad, Vice President (retired), Technology Strategy, Xerox Corporation (resigned January 2008; deceased September 2009)
Catharine R. Stimpson, Dean, Graduate School of Arts and Science, and University Professor, New York University
Richard P. Wheeler, Vice Provost, University of Illinois at Urbana-Champaign

Staff
Charlotte V. Kuh, Study Director
Peter H. Henderson, Senior Program Officer
James A. Voytuk, Senior Program Officer
John Sislin, Program Officer
Michelle Crosby-Nagy, Research Assistant
Kara Murphy, Research Assistant
Rae E. Allen, Administrative Coordinator
Sabrina Hall, Program Associate

Data Panel for the Assessment of Research-Doctorate Programs

Norman M. Bradburn, Ph.D., ***(Chair),*** Tiffany and Margaret Black Distinguished Service Professor and Provost *Emeritus*, University of Chicago
Richard Attiyeh, Ph.D., Vice Chancellor for Research, Dean of Graduate Studies, and Professor of Economics *Emeritus,* University of California, San Diego
Scott Bass, Ph.D., Provost, The American University
Julie Carpenter-Hubin, M.A., Director of Institutional Research and Planning, The Ohio State University
Janet L. Greger, Ph.D., Vice Provost for Strategic Planning, University of Connecticut *(retired)*
Dianne Horgan, Ph.D., Associate Dean of the Graduate School, University of Arizona
Marsha Kelman, M.B.A., Associate Vice President, Policy and Analysis, Office of the President, University of California
Karen Klomparens, Ph.D., Dean of the Graduate School, Michigan State University
Bernard F. Lentz, Ph.D., Vice Provost for Institutional Research, Drexel University
Harvey Waterman, Ph.D., Associate Dean for Academic Affairs, Graduate School-New Brunswick, Rutgers, The State University of New Jersey
Ami Zusman, Ph.D., Coordinator, Graduate Education Planning & Analysis, Office of the President, University of California *(retired)*

BOARD ON HIGHER EDUCATION AND WORKFORCE

Earl Lewis
Provost, Executive VP for Academic Affairs, and Professor of History
Emory University

Paula Stephan
Professor of Economics
Andrew Young School for Policy Studies
Georgia State University

Staff
Charlotte Kuh, Deputy Executive Director, Policy and Global Affairs
Peter Henderson, Director, Board on Higher Education and Workforce
James Voytuk, Senior Program Officer
Mark Regets, Senior Program Officer
Christopher Verhoff, Financial Associate
Michelle Crosby-Nagy, Research Associate
Sabrina Hall, Program Associate

PREFACE AND ACKNOWLEDGMENTS

This methodology was originally released in 2009 in advance of the Data-Based Assessment of Research-Doctorate Programs to educate those who will be using the assessment in the details of its construction. This revision is appearing at the same time as the assessment because the methodology has changed somewhat and users should have access to a guide that reflects those changes. Although it builds on earlier NRC assessments, the methodology of assessments has been significantly altered and the range of data used in constructing rankings has been considerably expanded. At the urging of members of the graduate community, the Committee to Assess Research Doctorate Programs has produced this guide because the methodology is statistically complex and it is important to know what to look for when the range of rankings for each program is released.

A data-based study as large as the one described in this methodology guide would be impossible without the time, care, and assistance of hundreds of people in universities who gathered and checked data for each of their programs to assist with the Assessment of Research Doctorate Programs. We will not name all these people, but the committee is extraordinarily grateful to them. The questionnaires were developed with the assistance of the committee's Data Panel, chaired by Norman Bradburn, whose membership appears following the list of the committee. We are also grateful to the staff of Mathematica Policy Research, our data contractor, which not only collected the data, but helped us with questionnaire wording, sampling plans, and model implementation. Geraldine Mooney and David Edson were the able leaders of a large MPR team.

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the National Academies' Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible, and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the process.

We wish to thank the following individuals for their review of the original version of this report: John Bailar, University of Chicago; John Burris, Burroughs Wellcome Fund; Michael Brick, Westat, Inc.; Joseph Cerny, University of California, Berkeley; Karen DePauw, Virginia Polytechnic Institute; Robeson Taj Frazier, University of California, Berkeley; Andrew Gelman, Columbia University; Claudia Goldin, Harvard University; Valen Johnson, M.D. Anderson Cancer Center; Sheryl Lightfoot, University of Minnesota; Daniel Mote, University of Maryland; Risa Palm, State University of New York; William Press, University of Texas; Raul Ramos, University of Houston; Stephen Stigler, University of Chicago; Dawn Terkla, Tufts University; Andrew Wachtel, Northwestern University; George Walker, Cleveland State University; John Wiley, University of Wisconsin; and Lilian Wu, International Business Machines Corporation.

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations, nor did they see the final draft of the report before its release. The review of this report was overseen by Stephen Fienberg, Carnegie Mellon University and Lyle Jones, University of North Carolina, Chapel Hill. Appointed by the National Academies, they were responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the authoring committee and the institution.

Finally, we wish to thank our funders, the National Science Foundation (OIA-0540823), the National Institutes of Health (N01-OD-4-2139, TO#170), the U.S. Department of Energy (DE-FG02-07ER35880), the Alfred P. Sloan Foundation (2004-3-20), the Andrew W. Mellon Foundation, and the President's Committee of the National Research Council; and the 212 universities that participated and contributed financial support to the assessment.

Contents

1

Introduction

The Data-Based Assessment of Research-Doctorate Programs conducted by the National Research Council (NRC) provides data that allow comparisons to be made among similar doctoral programs around the United States, with the goal of informing efforts to improve current practices in doctoral education. The assessment, which covers doctoral programs in 62 fields at 212 institutions[1], offers accessible data about program characteristics that will be of interest to policymakers, researchers, university administrators, and faculty, as well as to students who are considering doctoral study. Furthermore, the assessment analyzes and combines these data to create illustrations of two ranges of rankings, calculated in different ways, which allow the comparison of different doctoral programs within a field.

PURPOSE OF THE METHODOLOGY GUIDE

This methodology guide is intended mainly for those people in universities who are asked to explain the results of the NRC Data-Based Assessment to their presidents and provosts. This intended audience consists primarily of faculty, many of whom are serving as graduate deans and graduate program directors, as well as institutional researchers. Other potential audiences include those people asked to explain the use of the study to the public, as well as those students who are considering doctoral study. The *Guide* was first released in prepublication form in July 2009, in response to requests from the graduate school community that the NRC provide information on the study methodology in advance of the release of the assessment, so that users could prepare for it. Since July 2009, portions of the methodology used in the assessment have changed, and this volume has been updated to reflect those changes. The assessment itself is a separate document: a brief report on doctoral education in U.S. universities accompanied by tables that contain data, dimensional measures, and two illustrative ranges of rankings for programs on a field-by-field, program-by-program basis.[2]

[1] 221 institutions in all, when joint programs among universities are included. 59 fields with rankings plus 3 unranked fields.
[2] The original pre-publication version of this Guide, which appeared as a PDF in July 2009, explained the basis for using faculty values to develop one set of overall ratings and rankings that combined in a statistically valid way the two measures that are described here. It turned out, however, that the production of rankings from measures of

This methodology guide is organized into the following chapters:

- **A brief description of the data—**This section lays out how the study was designed and how the data were collected. In particular, it covers the recruitment of the participating institutions, the questionnaires, how the taxonomy of fields was determined, the determinants of program inclusion, the reasons for dropping some programs and some fields, how a sample survey of faculty was used in obtaining ratings[3], and how the faculty questionnaire was used to determine direct measures of quality.
- **How ratings in three dimensions are calculated—**In addition to the overall illustrative measures provided by the assessment for each program at each institution, dimensional measures were constructed in three areas: research activity, student support and outcomes, and student and faculty diversity. These measures take into account only the variables relevant to each area.
- **Calculating the overall illustrative rankings of a program—**This section covers the sources of variability in ratings, direct measurement of quality as perceived by faculty, regression-based measures of the importance of measured variables to program quality, and how ratings are calculated and converted to a range of rankings. The calculation includes all the variables (20 for non-humanities fields, 19 for the humanities and computer science).
- **An example—**The calculation of the ranges of rankings for a program in economics is presented and explained.

This guide also presents technical information about the current study. Appendix A describes the statistical techniques used to obtain the ratings and ranges of rankings and is intended for those interested in the statistical basis of the summary measures. Appendix B shows the range of rankings for the dimensional measures for 117 (anonymous) programs in economics as an example.
Additional technical data and background information can be found in the appendixes in *A Data-Based Assessment of Research-Doctorate Programs in the United States*. These include the questionnaires used to obtain the data about the universities, programs, faculty, and students (Appendix D); a detailed description of the 20 variables used in the calculations of the overall ranges of rankings (Appendix E); the R and S coefficients for each field (Appendix F); the correlation for median R and S rankings by broad field (Appendix G); and the average number of ratings obtained per program in the sample survey (Appendix H). Finally, a list of all the programs and their institutions by field can be found at http://sites.nationalacademies.org/pga/Resdoc/.

---

quantitative data was more complicated and had greater uncertainty than originally thought. As a consequence, the committee did not combine the two measures, and instead presented them as two illustrative rankings. Neither one is endorsed or recommended by the National Research Council (NRC) as an authoritative conclusion about the relative quality of doctoral programs.

[3] We use the term rating to mean a number on a scale from 1 to 6 that indicates the perceived quality of a program, or the statistically estimated perceived quality. Ratings from many raters were aggregated for programs as described in this guide and were then arranged in order, from highest to lowest, to yield a program ranking. A rating is a score. A ranking is calculated from an ordered list of ratings. In our study, we calculate multiple ratings for each program, and from the multiple ratings, obtain ranges of rankings.

Data for a Dynamic Discussion

The assessment has collected a great deal of data from doctoral programs across the United States, and it has statistically summarized these data along a variety of dimensions. The data that were assembled with great effort by U.S. research universities and their faculty, combined with the analytical talent of the many experts with whom the committee consulted, have enabled the production of a study with procedures designed to provide a richer array of results from those of previous NRC efforts and from those of commercial vendors

This study and its methodology, however, are merely the beginning of an informed discussion, not the last word. Users of the assessment and its methodology should understand that it was not the intent of the assessment committee to produce the final verdict (as of 2006) on the characteristics and quality of doctoral programs. Rather, we intend to present data that are relevant to the assessment of doctoral programs and to make them available to others. Users will want to bring to these data their own knowledge of programs and to compare the assessment that the NRC has produced with that knowledge. This should be a dynamic process that leads to further discussion and insights.

The committee seeks to make users aware of the strengths and limitations of the data and believe in the importance of this dynamic process. It has operated under the assumption that outstanding programs have certain measurable characteristics in common. For example, one can see evidence of a vibrant scholarly community by looking at measures of the number of faculty who produce scholarship and whose scholarship is recognized through citations, awards given by scholarly societies, and the percent of the faculty who receive grants. Nonetheless, the question of assessing how well a program accomplishes the dual objectives of conducting research and educating students to become scholars, researchers, and educators is a complex one.

The quality of doctoral programs is a multidimensional concept, and assessing that quality requires highlighting some of the more significant factors underlying it. This study has attempted to collect data that will capture this multidimensionality and to design measures that will best reflect it. Among the dimensions that we have sought to measure are: (1) the research activity of program faculty; (2) student support and outcomes; (3) diversity of the academic environment; and, taking these measures into account, (4) two summary measures that provide ranges of rankings of the estimated overall quality of programs, which includes all these separate dimensions, included with differing weights, and which are based on recent quantitative measurements. Each of these five measures necessarily collapses interesting and informative measures of doctoral programs. The committee hopes that users of the study will want to mine the data that underlie each metric, to examine additional information collected in the course of the study, and then construct their own comparisons. This will be possible by using the online spreadsheet that will accompany the final report.

In this undertaking, we were necessarily limited to examining what is countable[4]. Many will argue that program quality goes well beyond what can be measured: the existence of a scholarly community, the creative blending of interdisciplinary perspectives, or the excitement generated by the exploration of new paradigms. We agree. The committee also understands that some of these important qualitative dimensions will elude even the most carefully conceived quantitative measures. In order to capture as fully as possible those subjective dimensions that correlate with excellence in doctoral education, however, it surveyed a sample of program faculty about the perceived quality of a sample of programs in their individual fields and then used standard statistical techniques to find the measurable characteristics that best correlated with these subjective estimates of program quality. It used this technique to create regression-based, or R-based, ranges of rankings. It also asked faculty members in each field for their explicit views of the characteristics that are most important in facilitating a strong Ph.D. program. The weights derived from this survey were used to create survey-based, or S-based, ranges of rankings.

[4] "Perceived quality," a notion that underlies the rating part of the study, is measurable, but not countable. Most of the other variables in the study, such as numbers of faculty, students, citations, or publications are countable.

2

The Data and How They Were Obtained

The long history of the NRC Assessment of Research Doctorate Programs in the United States—this is the third in a series of such assessments since 1982—will not be recounted in detail here. Rather, we will offer a shortened history that begins with the decision of the National Research Council to undertake another study following the assessment published in 1995. The first step in the process of developing this new assessment was the publication of *Assessing Research Doctoral Programs: A Methodology Study* (“the Methodology Study“)[5], which was completed in 2003 and provided a roadmap for the large-scale study. At this point, universities still had to be recruited to join in the study, the final taxonomy of disciplines had to be settled, and the questionnaires had to be finalized and administered.

RECRUITING UNIVERSITIES

In November 2006 the chairman of the National Research Council, Ralph Cicerone, notified presidents and chancellors of U.S. universities offering doctoral degrees of the NRC’s intention to conduct a new assessment of doctoral programs. The universities were asked to contribute funding to the project, with the amount determined by a sliding scale that reflected the number of doctoral degrees in selected fields granted in 2003-2004 according to the National Science Foundation’s Survey of Doctoral Recipients.[6] Two hundred and twelve universities chose to participate.[7] Most of the data collection was carried out in late fall 2006 and spring 2007. Data were checked through fall 2007 via correspondence with many institutions. Data collection was completed in the spring of 2008. At this point the study had collected data for more than 5,000 programs in 62 fields in the physical sciences and mathematics, agricultural and life sciences, health sciences, engineering, social sciences, and arts and humanities.[8] Unless otherwise stated,

---

[5] National Research Council, *Assessing Research-Doctorate Programs: A Methodology Study* (Washington, D.C.: National Academies Press, 2003

[6] A contribution was not required for participation, but almost all of the participating universities did contribute funds.

[7] The institutions that chose not to participate generally had very few doctoral programs and often were undergoing administrative reorganization. Although the NRC followed up with institutions that did not respond, a handful of institutions that had been invited were excluded because of non-response. The total number of administrative units that participated is 221 and includes 9 combinations of universities that offer programs together.

[8] Data were collected for 67 fields in all, but 6 of these were emerging fields with too few programs to rate. Only partial data were collected for 5 of these fields. The other fields that were not rated were: Languages, Societies, and

the data reported in this study are for the 2005-2006 academic year. The universities and their programs are listed at: http://sites.nationalacademies.org/pga/Resdoc/.

THE TAXONOMY

At the same time as the universities were being recruited, the committee consulted widely in order to settle on a taxonomy of disciplines.[9] To assist in this task, it examined the taxonomy of fields used by the National Science Foundation (NSF) in its Doctorate Records File,[10] reviewed the classification of instructional programs (CIPS) of the U.S. Department of Education, and consulted with a number of scholarly societies. These societies were especially helpful when it came to the life sciences, because the taxonomy used in the 1995 NRC study for that area had become outdated. In particular, interdisciplinary study in the life sciences had grown considerably since 1995. This is reflected in the current study by the addition of an interdisciplinary field, "Biology/Integrated Biology/Integrated Biomedical Sciences," which includes 120 programs. Most of the other changes from the 1995 NRC study served to expand the disciplines that were included. For example, programs in agricultural fields, public health, nursing, public administration, and communication were added. The committee decided not to include doctoral programs in schools of education, because in many cases, research and practice-oriented doctoral programs could not be separated. A separate study of these programs is being conducted under the auspices of the American Education Research Association.

The criteria for inclusion of a field or a discipline in the study were that it had produced at least 500 Ph.D.'s in the five years prior to 2004-2005, and that there were programs in the field in at least 25 universities.[11] The criterion for inclusion of a program was that it had produced at least five Ph.D.'s in the five years prior to 2005-2006.[12] Given these criteria, each university chose which of their programs to include. The disciplines and programs covered by the study can be found at: http://sites.nationalacademies.org/pgs/Resdoc/.

QUESTIONNAIRE CONSTRUCTION AND DATA COLLECTION

During the winter of 2005-2006, a panel consisting of graduate deans and institutional researchers met to review the questionnaires that had been developed for the methodology study and to suggest additional and alternative questions. Once the draft questionnaires had been posted on the project Web site, many suggestions were also received from the universities. The

---

Cultures, computer engineering, and engineering science and materials, all of which had fewer than the required 25 programs.

[9] A provisional taxonomy had been suggested in *Assessing Research-Doctorate Programs: A Methodology Study.* This was revisited by a panel of the current Committee.

[10] The Doctorate Records File, administered by the National Science Foundation (NSF), is a joint data gathering activity of NSF, the U.S. Department of Agriculture, U.S. Department of Education, U.S. Department of Energy, the National Institutes of Health, and the National Endowment for the Humanities

[11] The fields of German and classics were included, although they did not meet these criteria, because they had been included in earlier NRC assessments. In 2006, not only were they included for historical reasons, but they qualified on the basis of the number of programs in the field.

[12] The dates for the test of field inclusion differ from those for program inclusion because of the lag in NSF data on Ph.D. production by field. Program data, which were obtained from the universities, were more current.

questionnaires were finalized in November 2006 and they appear in Appendix D of *A Data-Based Assessment of Research-Doctorate Programs*. The administration of the questionnaires involved the following steps:

- **Questionnaire design—**Five questionnaires were designed:
  1) an **institutional questionnaire**, which contained questions about institution-wide practices and asked for a list of doctoral programs at the institution.
  2) a **program questionnaire**, which was sent to each doctoral program in most cases[13]. In addition to questions about students, faculty, and characteristics of the program, programs were asked to provide lists of their doctoral faculty, and for five fields, their advanced doctoral students (see below)
  3) the **faculty questionnaire,** which asked individual faculty members about their educational and work history, grants, publications, what characteristics they felt were important to the quality of a doctoral program, and whether they would be willing to answer a survey asking them to provide ratings for programs in their field.
  4) the **student questionnaire**, sent to advanced students in English, chemical engineering, economics, physics, and neuroscience, which asked about student educational background, research experiences while in the program, program practices that they had experienced, and post-graduation plans.
  5) the **rating questionnaire**, which was sent to a stratified sample of those who had answered on the faculty questionnaire that they were willing to provide ratings of programs in their field.

  The operation of administering all these questionnaires was conducted by the committee's contractor, Mathematica Policy Research, in close collaboration with NRC staff. All questionnaires were submitted and approved by the Institutional Review Board (IRB) of the National Research Council and most institutions also received approval from their own IRBs.

• **Data Collection—**Each of the participating universities was asked to name an institutional coordinator (IC) who would be responsible for collection of data from the university. On the institutional questionnaire, the IC provided the names of the programs at that university that met the NRC criterion for inclusion. Each of these programs was then sent the program questionnaire through the IC. Some universities had a well-developed centralized data-collection capability and provided much of the data centrally. Others did not and gave the program questionnaires to each of their programs to complete. Each program was asked for a list of faculty members who were involved in doctoral education according to the NRC definition of a program that was given on the institutional and program questionnaires. On the program questionnaire, the committee asked respondents to divide their program faculty into three groups: (1) core faculty, who either were actively supervising doctoral dissertations or serving on an admissions or curriculum committee for the doctoral program; (2) new faculty, who were tenured, or tenure-track faculty, who had been hired in the previous three years and were expected to become core faculty; and (3) associated faculty, who were not core faculty in the program, but were working in the

[13] Some large institutions with well-equipped institutional research offices answered those program questions they could centrally and then sent the remaining questions to the doctoral programs to answer.

program supervising dissertations and were regular faculty members at the institution. The faculty questionnaire was then sent to core and new faculty in each program and included a section (Section G) asking what aspects of doctoral programs the faculty member thought were important to quality.

The committee requested that programs in five fields (physics, English, chemical engineering, economics, and neuroscience) provide lists of enrolled students who had been admitted to candidacy. These students were then each sent a copy of the student questionnaire. All questionnaires were delivered and answered online. Selected results of the student survey are provided in the final report, but are not discussed in this guide. As part of the faculty questionnaire, faculty members were asked if they would be willing to complete a rating survey. Those who indicated they were willing were put into a pool that was used to obtain the stratified sample of raters for the rating survey. Although response rates varied by field, there were no detectable characteristics of non-respondents that would suggest response bias.

• **Sampling for the rating survey—**Programs and raters within a field were classified according to the size of the program (measured by faculty size) and the program's geographic region. Raters were also classified by faculty rank. In the fields with a large number of programs, 50 programs were sampled at random from a stratified classification. In fields with a smaller number of programs, 30 programs were chosen in a similar manner. A sample of raters in each field was chosen so that the sample duplicated the distribution by program size, faculty rank, and geographic region for all programs in the field. Each rater was given a set of 15 programs to rate on a six-point scale, for which 1 was "not adequate for doctoral education" and 6 was "distinguished." The questionnaire also asked the rater's familiarity with each program and provided information about the program and a reference to the program Web site. On average, programs received ratings from about 58 percent of the selected raters who had been given data about them. Non-respondents were replaced by other raters from the same stratum until almost every program had been rated by 40 raters.[14] The numbers of raters for programs in each rated field are shown in Appendix H of *A Data-Based Assessment of Research-Doctorate Programs* [15].

• **Method of collecting publications, citations, and awards—**With the exception of fields in the humanities, publications and citations were collected through the Institute for Scientific Information (ISI), now a part of Thomson Scientific, and matched to faculty lists for fields in the sciences (including the social sciences). To assist in matching publications to faculty, faculty were asked for a list of ZIP codes that had appeared on their publications. These were used to match publications to faculty who had moved and to distinguish among faculty with the same name and field. Although faculty were also asked about their publications in Section D of the faculty questionnaire, these lists were used only to check the completeness of the

[14] The average number of raters taken over all programs was 44. See Appendix H of *A Data-Based Assessment of Research-Doctorate Programs*. Since the committee did not know in advance how many programs there would be in each discipline, special treatment was given during the regression calculations to programs in disciplines with fewer than 35 programs. These were combined with another field that had similar "direct" weights in order to obtain the regression-derived ratings

[15] Languages, Societies, and Cultures was a special case that was not rated when it became clear to the committee that the programs included in the "field" were too heterogeneous for ratings to be obtained that were comparable across the field and that no subfield had more than 20 programs. Respondents included programs in Italian, romance languages, Russian studies, Middle Eastern studies, African studies, and a number of other fields. Full data about these programs can be found in the spreadsheet accompanying the final report. Computer engineering and Engineering Science and Materials also had fewer than 25 programs and as a result were not rated.

Thomson-Reuters/ISI data. The citation count is for the years 2000-2006 and relates to papers published between 1981 and 2006. In the case of the humanities, for which we do not have a comprehensive bibliographic source, we analyzed faculty members' curriculum vitae*,* which were submitted along with the faculty questionnaire or the list they provided in answer to the questionnaire. We then counted books and publications going back to 1996 and recorded these counts, giving books a weight of 5 and articles a weight of 1. In the case of computer science, papers from refereed conferences are also an important form of scholarly activity. Counts of these papers, taken from faculty curriculum vitae, were added to the Thomson-Reuters indexed counts of articles for this field. Finally, lists of honors and awards were collected from 224 scholarly societies for all fields and differentiated between "highly prestigious" awards, which received a weight of 5, and other awards, which received a weight of 1.

• **Key variables—**Twenty-one key variables[16] were identified by the committee for inclusion in the rating process; these are described in Appendix E of *A Data-Based Assessment of Research-Doctorate Programs*. One variable that the committee wished to include—the number of student publications and presentations—was excluded because of lack of data. Most of these variables are expressed as *per capita* or "intensive" variables; that is, the committee divided the measure of interest (e.g., publications, citations) by the "allocated" faculty in the program, or, in the case of citations, we divided citations by the number of publications for each faculty member. This allocation was designed to assure that no more than 100 per cent of a faculty member was assigned to all programs taken together. The use of these key variables is described in Chapter 3.[17]

• **Final data review**—Once all the data had been collected, they were reviewed by NRC staff for completeness and consistency. The institutional coordinators were asked to revise anomalous data and populate missing cells. If, after this request, the programs were still unable to provide missing data, two procedures were followed: If data on two or fewer measures were missing, the cells were populated with the mean value for programs that had provided data.[18] If

[16] There were only 19 for the humanities fields and computer science, since citation data were unavailable for the publications obtained from c.v.'s.

[17] The justification of each of the variables is discussed in the final report. Two variables, however—one controversial and one novel—should be mentioned at this point. There is a large literature about the use of citations as a measure of excellence. A citation measure for an individual faculty member may be manipulated by self-citation. Flawed results may be highly cited but not indicative of quality. We grant the validity of these objections, but remind the reader that we are aggregating citations across the publications of all the faculty members in a program. Considering aggregated data, within a field, subdisciplines can have varying patterns of productivity, and the number of citations an article may receive are not independent of the size of the sub-discipline, so that the value of the measure for a program will depend on its specialty composition, not the quality of the program. The final report has a short discussion of these pitfalls. We use the variable here, in intensive form, because other things equal, we believe that a program whose faculty are more cited and that has a greater number of citations per publication will be a higher-quality program. The novel variable is interdisciplinarity. It, too, is discussed at greater length in the final report. The committee measured interdisciplinarity by the percent of program faculty who are serving on dissertation committees from outside the program (associated faculty). This is an imperfect measure, since it will depend on institutional practices; e.g., how broad doctoral programs are. It felt, however, that some measure, however imperfect, would be informative. This variable rarely shows up as important in determining program ratings.

[18] These values are identified in the data tables that accompany the final report by being printed in italics. Eight hundred fifty-four programs out of 4,915 total had at least one missing value. Programs were dropped if they did not submit a faculty list, so there were no missing values for the publications, citations, or awards measures.

data for three or more measures were missing, the program was dropped and the institutional coordinator was informed. If the data were then provided, the program was reinstated. Program names and assignment to a field were also reviewed by staff, and the institutional coordinator was consulted and if anomalies were found the recommendation by the IC for field assignment was followed.

3

Ratings in Specific Dimensions: The Dimensional Measures

The dimensional measures are provided to assure that measures of a broad range of characteristics of doctoral programs are available. They are divided into three categories: (1) research activity, (2) student support and outcomes, and (3) diversity of the academic environment. Each of the dimensional measures begins with the measures relating to one dimension of doctoral program performance, applies the weights from the faculty survey about what program characteristics contribute to quality, and then constructs a range[19] of rankings for each program based on this dimension of the data, taking into account variability in the data and in the choice of raters. They are dimensional in the sense that they provide more focused measures than the overall measures of ranges of rankings, but they are central to the calculation of those ranges.

Some specifics about the calculation of these measures follow.

• **How the weights are obtained**—As part of the NRC faculty questionnaire, the committee asked faculty to indicate the relative importance of different characteristics of doctoral programs; this was done through the multipart question that makes up Section G of the faculty questionnaire (see Appendix D of *A Data-Based Assessment of Research-Doctorate Programs*). Faculty were questioned about faculty quality, student characteristics, and program characteristics. First they were asked to indicate up to four characteristics in each category that they thought were important to program quality. Each characteristic that was listed received an initial score of 1. These preferences were then narrowed by asking the faculty members to identify a maximum of **two** characteristics in each category that they thought were most important. These characteristics each received a score of 2. A final question asked faculty members to indicate the relative importance of each category by assigning category weights whose values summed to 100. For each individual faculty member, the weight for a variable was calculated as the sum of the "votes" that it received times the importance assigned to the category that contained it. The weight for a variable in a discipline was the average weight taken across all faculty members in it. We took into account variability in raters' opinions and uncertainties due to missing data and the fact that

[19] When we use the term "range," we are referring to a range that covers 90 percent of the rankings for a program. This is obtained by eliminating the highest and lowest 5 per cent of the rankings.

some measures were sampled at one point in time.[20] Approximately 86 percent of the faculty responded. Their responses permitted calculation of the set of "direct" or survey-based weights. Although there was some variation in the faculty responses, they were generally in agreement that publications and citations were the most important factors in program quality[21]. Every variable, however, received some weight[22]. These weights were used to construct the dimensional measures. The average weights for programs in each broad field are shown in Appendix E of *A Data-Based Assessment of Research-Doctorate Programs*, and an example of ranges of rankings for programs in economics is shown in Appendix B.

• **Research activity—**This dimensional measure relates to various ways to gauge the contribution of research: publications, citations (except for the humanities and computer science), the percent of the faculty holding research grants, and recognition of scholarship as evidenced by honors and awards. Specifically, the components of the research activity dimensional measure are: average publications per allocated faculty member,[23] average citations per publication, percent of core and new doctoral faculty respondents holding grants, and awards per allocated faculty member. Publishing patterns and the availability of research funding and awards for scholarship vary by field, but the weight placed on publications per faculty member is remarkably consistent—about 30 percent—across fields. Research activity is the dimensional measure that most closely tracks the overall measures of program quality, because in all fields, both the survey-based or direct measure—based on abstract faculty preferences—and the regression-based measure also puts high weight on the measures of research productivity in addition to the measure of program size.

• **Student support and outcomes—**This measure combines data on the percent of students fully funded in the first year, the percent of students completing their degrees in a given time period, time to degree, placement in academic positions (including academic postdoctoral positions), and whether a program collects data about the employment outcomes for its students. The committee found that faculty typically placed a larger weight on student support and completion rates than on median time to degree, academic placement, or whether a program follows the employment outcomes of its students.[24] There is surprising uniformity

[20]There is some uncertainty in the values of the program variable values themselves. Some of the 20 program variables used to calculate the ratings also vary or have an error associated with their values due to year-to-year fluctuations. Data for five of the variables (publications per faculty, citations per publications, GRE scores, Ph.D. completion, and number of Ph.D.'s) were collected over time, and averages over a number of years were used as the values of these program variables. If a different time period had been used, the values would have been different. To express this type of uncertainty, a relative error term, e_{jk}, was associated with each variable value. For details, see Appendix A.

[21] Since citation data were unavailable for the humanities fields and for computer science, the contribution of citations to the rating for these fields had to be set to zero, an important omission. The effect is to lower the rating for those programs whose faculty have relatively but few but highly influential publications from what it would have been had citations been included.

[22] All "direct" weights are used in the calculation of the Dimensional Measures.

[23] Because many faculty members supervise dissertations in more than one program, faculty members were allocated across these programs so that the total, taken across all programs, equaled one or less (in the case in which the faculty member was in a professional school).

[24] Ideally, we would have used a measure such as employment in one's field 5 years after receipt of Ph.D., but many programs did not collect such data. The committee hoped that including this measure would encourage more programs to pay attention to post-degree outcomes for their graduates.

across broad fields on the weights, which are shown in Appendix E of *A Data-Based Assessment of Research-Doctorate Programs*.

• **Diversity of the academic environment**—The diversity measures did not appear as major factors in determining the overall perceived quality of programs. Taken separately, there are definite patterns for variables that faculty thought were more important, and these vary by field. The measures that are included in this dimensional measure are: the percent of faculty and percent of students who are from underrepresented minority groups, the percent of faculty and the percent of students who are female, and the percent of students who are international (that is, in the United States on a temporary visa). In terms of field differences, most fields place the highest weight on the percentage of students from underrepresented minority groups. In the health sciences, social sciences, and humanities, relatively high weights are also placed on the percentage of faculty who are underrepresented minorities. The percentage of international students was not highly weighted, except for the physical sciences. These weights, by broad field, are shown in Appendix E of *A Data-Based Assessment of Research-Doctorate Programs*.

What is interesting about the dimensional ratings is that, with the exception of the research activity measure, they produce program rankings that are quite different from the overall ratings. This can be seen for one field in the table in Appendix B. Excellence in doctoral programs is not uni-dimensional. Some students may prefer a program where they can be assured of steady funding and a short time to degree, even if it is not a program that is perceived as stellar in terms of the productivity of its faculty. Similarly, a program that is more diverse may be preferable to many students, although diversity bears only a tenuous relation with the usual measures of scholarly productivity. Users of the assessment should be aware of these different dimensions, because each presents the characteristics of an individual doctoral program from a different perspective.

4

Two Illustrations of Overall Ratings of Program Quality

The dimensional measures provide a summary of program performance along individual dimensions that are of importance in doctoral education. The illustrations of overall ratings described below combine the variables that make up the dimensional measures into two measures, one using the weights from the regression, and one using the weights from the survey. These two measures are illustrative of ways of deriving weights that can then be applied to program data to obtain different views of program quality. Other sets of weights can, of course, be chosen, which is why we view them as illustrative. This section describes in non-technical terms how two overall ratings for a program are calculated. Readers who wish more technical detail are referred to Appendix A.

THE OVERARCHING IDEA

There is a great deal of uncertainty in the ratings of the quality of programs. Uncertainty can come from a variety of sources. For example, although many academics may think that they can identify the top five or ten programs in their field, this certainty about perceived quality decreases as more and more programs are included. Furthermore, one program may be strong in one area while a second program's strengths may lie in a different area. Faculty asked to rate programs may differ in their views about the importance of these strengths, and the programs may differ in various characteristics, many of which may be considered important to the perceived quality of a doctoral program.

Describing this uncertainty was a key task of the predecessor committee that produced *Assessing Research-Doctorate Programs: A Methodology Study*.[25] This committee examined the

[25] National Research Council., *Assessing Research-Doctorate Programs: A Methodology Study*. Washington, D.C. 2003.

methodology of the 1995 study and recommended that the next study rely more explicitly on program data. It also contained two key recommendations as to how the methodology of obtaining reputation measures should be revised:

> "The next study should have sufficient resources to collect and analyze auxiliary information from peer raters and the programs being rated to give meaning and context to the rating ranges that are obtained for the programs…." (p. 5)
>
> and
>
> "Re-sampling methods should be applied to ratings to give ranges of rankings for each program that reflect the variability of ratings by peer raters. The panel investigated two related methods, one based on Bootstrap re-sampling and another closely related method based on Random Halves, and found that either method would be appropriate." (p. 5)

The dimensional ratings, described in the previous section, fulfill the first recommendation. This section describes how the second recommendation was followed and combined with the first to obtain two illustrative overall ratings which were arranged from highest to lowest to produce a range of rankings for each program within a field.

THE OVERALL APPROACHES

A schematic description of the overall approaches appears in Box 4-1 and is described in the text:

Box 4-1

Faculty

Students

Institutions and Programs

Existing Data

1. DATA

- Answers to questions provided by 4,838 doctoral programs at 221 institutions and combinations of institutions in 59 fields across the sciences, engineering, social sciences, arts, and humanities covering institutional practices, program characteristics, and faculty and student demographics obtained through a combination of original surveys and existing data sources (NSF surveys and Thomson-Reuters publication and citation data).

2. WEIGHTS

- In two surveys shown in Appendix D of *A Data-Based Assessment of Research-Doctorate Programs* , program faculty provided the NRC with information on what they value most in Ph.D. programs:
- (1) Faculty were asked *directly* how important they felt 21 items in a list of program characteristics were (for S weights).
- (2) A sample of faculty rated a sample of programs in their field. These ratings were then related through regressions to the same items as appeared in (1) using a principal components transformation to correct for colinearity (for R weights).

3. ANALYSIS

- "Survey (S)" and "regression-based (R)" weights provided by faculty were used to calculate separate ratings, reflecting the multidimensional views faculty hold about factors contributing to the quality of doctoral programs.

4. RANGES OF RANKINGS

- Each program's rating was calculated 500 times by randomly selecting half of the raters from the faculty sample in Step #2 and also incorporating statistical and measurement variability. Similarly, 500 samples of survey based weights were selected.
- The R-weights and the S-weights were then applied to 500 randomly selected sets of program data to produce two sets of ratings for each program.
- These ratings for each of the 500 samples determined R and S rank orderings of the programs.
- A "range of rankings" was then constructed showing the *middle 90 percent* range of calculated rankings using the R and the S rankings separately. What may be compared, among programs in a field, are these ranges of rankings.

Faculty were surveyed to obtain their views on the importance of different characteristics of programs as measures of quality. Ratings were based on faculty members' views of how those measures related to program quality, as discussed in the chapter on dimensional measures. The views were related to program quality using two distinct methods: (1) directly, through answers to questions on the faculty survey; and (2) regression-based, obtained by asking faculty raters to provide program ratings for a sample of programs in a field and then relating these ratings, through a regression model that corrected for correlation among the characteristics, to data on the program characteristics. The two methods approach the ratings from different perspectives. The direct approach is a "bottom-up" approach that builds up the ratings from the importance that faculty members gave to specific program characteristics independent of reference to any actual program. This is the survey-based (S) rating. The regression-based (R) method is a "top-down" approach that starts with ratings of actual programs and uses statistical techniques to infer the weights given by the raters to specific program characteristics. The survey-based approach is idealized. It asks about the characteristics that faculty feel contribute to quality of doctoral programs without reference to any particular program. The first approach presented the respondent with 15 programs in his or her field and asked for ratings of program quality[26], but the responders were not explicitly queried about the basis of their ratings.

As is explained below, each rating is recalculated 500 times using different samples of raters. The program ratings obtained from all these calculations can then be arranged in rank order and, in conjunction with all the ratings from all the other programs in the field, used to determine a range of possible rankings.

Because of the various sources of uncertainty, which are discussed at greater length in Appendix A, each ranking is expressed as a range of values. These ranges were obtained by taking into account the different sources of uncertainty in these ratings (statistical variability from the estimation, program data variability, and variability among raters). The measure of uncertainty is expressed by reporting the end points of a range that includes 90 percent of all the ratings for a program. These are the 5th percentile point and the 95th percentile point. An example of the derivation of rankings for a program is given in Chapter 5.

In summary, we obtain two ranges of rankings for each program in a given field by first obtaining two sets of weights through two different methods, survey-based and regression-based. We then standardize all the measures to put them on the same scale and obtain ratings by multiplying the value of the standardized measure by the weights obtained from each method.

---

[26] The question given raters about program quality was:
The question given raters about program quality was:

On a scale from 1 to 6, where 1 equals not adequate for doctoral education and 6 equals a distinguished program, how would you rate this program?

Not Adequate for Doctoral Education	Marginal	Adequate	Good	Strong	Distinguished	Don't Know Well Enough
1	2	3	4	5	6	9

1 2 3 4 5 6 9

We obtain both the survey-based weights and coefficients from regressions through calculations carried out 500 times, each time with a different randomly chosen set of faculty, to generate a distribution of ratings that reflects their uncertainties. For both the S and the R rankings, we obtain the range of rankings for each program by trimming the bottom five percent and the top five percent of the 500 rankings to obtain the range that includes 90 percent of the program's rankings. This method of calculating ratings and rankings takes into account variability in rater assessment of what contributes to program quality within a field, variability in values of the measures for a particular program, and the range of error in the statistical estimation. It is important to note that these techniques give us a range of rankings for most programs. We do not know the exact ranking for each program, and to try to obtain one—by averaging, for example—could be misleading, because we have not imposed any particular distribution on the range of rankings.[27] The spreadsheet that presents the range of rankings for each program lists the programs alphabetically and gives the range for each program. Users are encouraged to look at groups of programs that are in the same range as their own programs, as well as programs whose ranges are above or below, in trying to answer the question, "Where do we stand?"

The next section provides an example of how the ranges of rankings were calculated for a particular program.

[27] For example, most of the rank ordered ratings could be at the top of the range.

5

An Example

The online spreadsheet accompanying the final report contains all the data for the 20 variables used in the rankings as well as data about additional variables, such as enrollment, and size, rank and tenure status of faculty. It also displays the ranges of rankings for the S-, R- and dimensional measures for each of the ranked programs in the study The spreadsheet may be found at http://www.nap.edu/rdp. Directions for launching and using the spreadsheet appear on worksheets with tabs labeled "Start" and "Guide." The "Master" spreadsheet allows for selection of programs by broad field and then by fine field. It is also possible to extract all the programs for one institution or selected institutions by clicking on the Sort button in the lower right corner of the Institution column heading cell, unchecking "select all" and checking the box beside the institution(s) of interest. Alternatively, it is possible to sort by field and to select programs of interest within a field using the sorting icon in the header of the Institution column. Demonstrations of common uses are available on the website.

The detail on the data that went into the rankings can be found by double-clicking on the program name or on any of the ranking values in the spreadsheet. If a value in a cell is displayed in green, it indicates that an analytic table giving background information is accessible by double-clicking within that cell. The program name is linked to a table that contains the detail for the individual program and for the field as a whole. Double-clicking individual ranking values leads to tables that provide the detail of the calculations that produced the 5th or 95th percentile rankings, respectively. In addition to double-clicking on individual cells containing values displayed in green, the user can access all the analytic tables pertaining to a program by highlighting any cell in the row of data for that program and then using the mouse or pointer to select any of the 11 table tabs at the bottom of the spreadsheet.

The first table tab (labeled "variables") shows the following: (1) the values that the program submitted or that were calculated from their data for each of the 20 variables with their corresponding standardized values (the online sheet is labeled "variables") as well as pairs of R- and S-coefficients (plus and minus one standard deviation from the average value) used in weighting the variables, and (2) the actual value and standardized program values and the actual coefficients that were used to calculate the rating corresponding to each endpoint of the 90 percent range of rankings for that program for each of the measures (Tables 2a through 6b, where *a* is the table showing the calculation of the 5th percentile value and *b* is the table showing the calculation of the 95th percentile value for the measure). Examples of these tables for an economics program are presented and discussed below.

Table 5-1 shows the values submitted by a program in economics and the range of R- and S- coefficients for the entire field. Columns 1 and 2 name and label the variables. Column 3 gives the program value for each of the 20 variables used in the overall rating (see Appendix E of *A Data-Based Assessment of Research-Doctorate Programs* for a description of these variables). Column 4 presents the standardized value of each variable in column 3. Thus, the relative strengths and weaknesses of a program (in terms of these 20 variables) can be seen by comparing the standardized values in column 4. Standardization involves taking the data values for each variable for all the programs in the field and putting them on a scale that has mean 0 and variance 1. This makes it easy to tell where a program stands relative to other programs in the field. In the example, this program is considerably above average on publications per faculty member (the standardized value is 2.175), but below the average on citations per publication (its value is -.322). Columns 5 through 6 give the one standard deviation range of the R- coefficients (weights) calculated for each variable used in rating all economics programs.[28] Columns 7 and 8 give the same sort of range for the S-coefficients. In economics, based on the R-coefficients, variables V2, V8, and V14 (cites per publication, GRE-Q and average number of Ph.D.'s) were assigned the largest weights. The S-coefficients are not the same as the R. In particular, the S-weights are high for publications per allocated faculty member and also for citations. GRE-Q and average Ph.D.s have considerably smaller weights than they do in the R-measure. The effect of the different weights can be seen in the detail of the rankings.

Tables 5-2a and 5-2b show the calculation for the 5th and 95th percentile for the regression-based rankings for this economics program. For the R measure, this program ranks from 53rd to 86th out of 117 programs. Tables 5-3a and 5-3b show a similar calculation for the S ranking. The program ranks between 27th and 53rd. The higher ranking is the result of the greater weight that the S-measure puts on publications per faculty member. It is worth noting that the standardized value of the program variables with variation, shown in column 4 of these tables, is not the same from one table to the next. This is because the values are selected randomly from within a range of variation, as described in Chapter 4.

[28] Five hundred regressions are run using half of the raters each time and 500 draws are made from randomly selected halves of the pool of direct ratings in order to construct the combined coefficients. The values presented show the range encompassed by plus or minus one standard deviation for each coefficient. See Appendix A for details.

TABLE 5- 1: Standardized Program Values and Range of Regression-Based (R) and Survey-Based (S) Coefficients

Institution Name: IHE University
Program: Economics
Program ID: 12345

(Col 1)	(Col 2)	(Col 3)	(Col 4)	(Col 5)		(Col 6)	(Col 7)		(Col 8)
		Program	Program Value	Regression-Based Coefficients			Survey-Based Coefficients		
Description	Variable	Value*	Standardized**	Minus 1 SD***		Plus 1 SD***	Minus 1 SD***		Plus 1 SD***
Publications per Allocated Faculty	V1	1.067	2.175	-0.017	to	0.033	0.202	to	0.206
Cites per Publication	V2	0.864	-0.322	0.182	to	0.228	0.200	to	0.204
Percent Faculty with Grants	V3	25.47%	-0.582	0.009	to	0.037	0.076	to	0.080
Percent Faculty Interdisciplinary	V4	5.88%	-0.641	-0.056	to	-0.026	0.017	to	0.019
Percent Non-Asian Minority Faculty	V5	7.69%	0.547	0.004	to	0.050	0.005	to	0.006
Percent Female Faculty	V6	12.50%	-0.440	-0.077	to	-0.020	0.006	to	0.008
Awards per Allocated Faculty	V7	0.000	-0.546	0.020	to	0.061	0.091	to	0.095
Average GRE-Q	V8	746	-0.269	0.119	to	0.158	0.081	to	0.084
Percent 1st yr. Students with Full Support	V9	100.00%	0.979	0.002	to	0.035	0.056	to	0.059
Percent 1st yr. Students with External Funding	V10	0.00%	-0.544	-0.030	to	-0.002	0.029	to	0.031
Percent Non-Asian Minority Students	V11	10.00%	0.069	-0.011	to	0.014	0.008	to	0.009
Percent Female Students	V12	44.44%	0.678	-0.064	to	-0.036	0.006	to	0.007
Percent International Students	V13	53.33%	-0.509	-0.033	to	0.002	0.015	to	0.017
Average PhDs 2002 to 2006	V14	5.4	-0.355	0.120	to	0.177	0.021	to	0.023
Percent Completing within 6 Years	V15	27.62%	-0.638	-0.043	to	-0.003	0.036	to	0.038
Time to Degree Full and Part Time	V16	5.67	0.232	-0.076	to	-0.042	-0.032	to	-0.030
Percent Students in Academic Positions	V17	11.11%	-1.405	-0.018	to	0.007	0.076	to	0.078
Student Work Space	V18	1	1.000	-0.056	to	-0.026	0.005	to	0.006
Health Insurance	V19	1	1.000	-0.040	to	-0.006	0.001	to	0.002
Number of Student Activities Offered	V20	17	0.439	0.003	to	0.035	0.017	to	0.019

* Col 3 shows data submitted by the program or calculated from these data.
** Col 4 shows standardized across all program values in the field, with mean of 0 and variance of 1.
*** Col 5 shows Minus 1 Standard Deviation from the Mean for the regression-based coefficients for the field as a whole
*** Col 6 shows Plus 1 Standard Deviation from the Mean for the regression-based coefficients for the field as a whole
*** Col 7 shows Minus 1 Standard Deviation from the Mean for the survey-based coefficients for the field as a whole
*** Col 8 shows Plus 1 Standard Deviation from the Mean for the survey-based coefficients for the field as a whole

TABLE 5-2a: 5th Percentile for the Regression-Based (R) Ranking Calculation

Institution Name: IHE University
Program: Economics
Program ID: 12345

(Col 1) Description	(Col 2) Variable	(Col 3) Program Value*	(Col 4) Standardized Program Value with Variation†	(Col 5) Regression Coefficient‡	(Col 6) Product of Col 4 X Col 5
Publications per Allocated Faculty	V1	1.067	2.113	0.024	0.051
Cites per Publication	V2	0.864	-0.444	0.213	-0.094
Percent Faculty with Grants	V3	25.47%	-0.618	0.054	-0.033
Percent Faculty Interdisciplinary	V4	5.88%	-0.605	-0.026	0.016
Percent Non-Asian Minority Faculty	V5	7.69%	0.430	-0.007	-0.003
Percent Female Faculty	V6	12.50%	-0.465	-0.025	0.012
Awards per Allocated Faculty	V7	0.000	-0.770	0.035	-0.027
Average GRE-Q	V8	746	-0.238	0.158	-0.037
Percent 1st yr. Students with Full Support	V9	100.00%	1.506	0.047	0.071
Percent 1st yr. Students with External Funding	V10	0.00%	-0.503	-0.002	0.001
Percent Non-Asian Minority Students	V11	10.00%	0.213	0.001	0.000
Percent Female Students	V12	44.44%	0.225	-0.078	-0.017
Percent International Students	V13	53.33%	-0.879	-0.050	0.044
Average PhDs 2002 to 2006	V14	5.4	-0.229	0.104	-0.024
Percent Completing within 6 Years	V15	27.62%	-0.842	-0.030	0.025
Time to Degree Full and Part Time	V16	5.67	-0.203	-0.049	0.010
Percent Students in Academic Positions	V17	11.11%	-1.340	-0.033	0.044
Student Work Space	V18	1	1.000	-0.044	-0.044
Health Insurance	V19	1	1.000	-0.013	-0.013
Number of Student Activities Offered	V20	17	0.184	0.009	0.002

Program Ranking: ***53 of 117 Programs***

* Col 3 is based on data submitted by the program or calculated from these data.
† Col 4 is standardized value for the set of perturbed program values that produced the 5th percentile ranking.
Standardized values have a mean of 0 and variance of 1.
‡ Col 5 is the regression-based weight for each variable

TABLE 5- 2b: 95th Percentile for the Regression-Based (R) Ranking Calculation

Institution Name: IHE University
Program: Economics
Program ID: 12345

(Col 1) Description	(Col 2) Variable	(Col 3) Program Value*	(Col 4) Standardized Program Value with Variation†	(Col 5) Regression-Based Coefficient‡	(Col 6) Product of Col 4 X Col 5
Publications per Allocated Faculty	V1	1.067	2.366	0.019	0.045
Cites per Publication	V2	0.864	-0.258	0.222	-0.057
Percent Faculty with Grants	V3	25.47%	-0.298	0.011	-0.003
Percent Faculty Interdisciplinary	V4	5.88%	-0.630	-0.055	0.034
Percent Non-Asian Minority Faculty	V5	7.69%	0.588	0.005	0.003
Percent Female Faculty	V6	12.50%	-0.314	-0.052	0.016
Awards per Allocated Faculty	V7	0.000	-0.640	0.026	-0.017
Average GRE-Q	V8	746	-0.286	0.146	-0.042
Percent 1st yr. Students with Full Support	V9	100.00%	0.761	0.008	0.006
Percent 1st yr. Students with External Funding	V10	0.00%	-0.843	-0.018	0.016
Percent Non-Asian Minority Students	V11	10.00%	0.069	0.011	0.001
Percent Female Students	V12	44.44%	1.297	-0.061	-0.079
Percent International Students	V13	53.33%	-0.340	-0.028	0.010
Average PhDs 2002 to 2006	V14	5.4	-0.688	0.125	-0.086
Percent Completing within 6 Years	V15	27.62%	-0.875	-0.034	0.029
Time to Degree Full and Part Time	V16	5.67	1.011	-0.083	-0.084
Percent Students in Academic Positions	V17	11.11%	-1.291	-0.005	0.006
Student Work Space	V18	1	1.000	-0.061	-0.061
Health Insurance	V19	1	1.000	-0.030	-0.030
Number of Student Activities Offered	V20	17	1.464	0.001	0.001

Program Ranking: ***86 of 117 Programs***

* Col 3 is based on data submitted by the program or calculated from these data.
† Col 4 is standardized value for the set of perturbed program values that produced the 95th percentile ranking.
Standardized values have a mean of 0 and variance of 1.
‡ Col 5 is the regression-based weights for each variable (See Appendix A).

TABLE 5- 3a: 5th Percentile for the Survey-Based (S) Ranking Calculation

Institution Name: IHE University
Program: Economics
Program ID: 12345

(Col 1) Description	(Col 2) Variable	(Col 3) Program Value*	(Col 4) Standardized Program Value with Variation†	(Col 5) Survey-Based Coefficient‡	(Col 6) Product of Col 4 X Col 5
Publications per Allocated Faculty	V1	1.067	2.585	0.205	0.529
Cites per Publication	V2	0.864	-0.360	0.206	-0.074
Percent Faculty with Grants	V3	25.47%	-0.642	0.077	-0.050
Percent Faculty Interdisciplinary	V4	5.88%	-0.578	0.017	-0.010
Percent Non-Asian Minority Faculty	V5	7.69%	0.488	0.006	0.003
Percent Female Faculty	V6	12.50%	-0.298	0.007	-0.002
Awards per Allocated Faculty	V7	0.000	-0.394	0.094	-0.037
Average GRE-Q	V8	746	-0.254	0.082	-0.021
Percent 1st yr. Students with Full Support	V9	100.00%	1.555	0.057	0.088
Percent 1st yr. Students with External Funding	V10	0.00%	-0.464	0.030	-0.014
Percent Non-Asian Minority Students	V11	10.00%	-0.042	0.008	0.000
Percent Female Students	V12	44.44%	0.612	0.007	0.004
Percent International Students	V13	53.33%	-0.608	0.016	-0.010
Average PhDs 2002 to 2006	V14	5.4	-0.275	0.021	-0.006
Percent Completing within 6 Years	V15	27.62%	-0.142	0.036	-0.005
Time to Degree Full and Part Time	V16	5.67	0.317	-0.031	-0.010
Percent Students in Academic Positions	V17	11.11%	-1.165	0.077	-0.090
Student Work Space	V18	1	1.000	0.006	0.006
Health Insurance	V19	1	1.000	0.002	0.002
Number of Student Activities Offered	V20	17	0.154	0.018	0.003

Program Ranking: ***27 of 117 Programs***

* Col 3 is based on data submitted by the program or calculated from these data.
† Col 4 is standardized value for the set of perturbed program values that produced the 5th percentile ranking.
Standardized values have a mean of 0 and variance of 1.
‡Col 5 is the survey-based weights for each variable

TABLE 5-3b: 95th Percentile for the Survey-Based (S) Ranking Calculation

Institution Name: IHE University
Program: Economics
Program ID: 12345

(Col 1)	(Col 2)	(Col 3)	(Col 4)	(Col 5)	(Col 6)
Description	**Variable**	**Program Value***	**Standardized Program Value with Variation†**	**Survey-Based Coefficient‡**	**Product of Col 4 X Col 5**
Publications per Allocated Faculty	V1	1.067	1.583	0.205	0.325
Cites per Publication	V2	0.864	-0.267	0.202	-0.054
Percent Faculty with Grants	V3	25.47%	-0.717	0.075	-0.054
Percent Faculty Interdisciplinary	V4	5.88%	-0.601	0.018	-0.011
Percent Non-Asian Minority Faculty	V5	7.69%	0.578	0.006	0.004
Percent Female Faculty	V6	12.50%	-0.298	0.007	-0.002
Awards per Allocated Faculty	V7	0.000	-0.745	0.092	-0.069
Average GRE-Q	V8	746	-0.286	0.083	-0.024
Percent 1st yr. Students with Full Support	V9	100.00%	0.727	0.057	0.041
Percent 1st yr. Students with External Funding	V10	0.00%	-0.683	0.030	-0.020
Percent Non-Asian Minority Students	V11	10.00%	0.011	0.009	0.000
Percent Female Students	V12	44.44%	1.016	0.007	0.007
Percent International Students	V13	53.33%	-0.834	0.016	-0.013
Average PhDs 2002 to 2006	V14	5.4	-0.523	0.021	-0.011
Percent Completing within 6 Years	V15	27.62%	-0.842	0.039	-0.033
Time to Degree Full and Part Time	V16	5.67	1.184	-0.031	-0.036
Percent Students in Academic Positions	V17	11.11%	-1.493	0.078	-0.117
Student Work Space	V18	1	1.000	0.006	0.006
Health Insurance	V19	1	1.000	0.001	0.001
Number of Student Activities Offered	V20	17	-0.068	0.018	-0.001

Program Ranking: ***54 of 117 Programs***

* Col 3 is based on data submitted by the program or calculated from these data.
† Col 4 is standardized value for the set of perturbed program values that produced the 95th percentile ranking.
Standardized values have a mean of 0 and variance of 1.
‡Col 5 is the survey-based weights for each variable

Tables 5-2a, 5-2b, 5-3a and 5-3b show the calculations of the 5th and 95th percentile rankings, respectively, for a particular program for the R and S measures. First, a set of randomly sampled regression coefficients or S-weights (depending on the measure) is used to obtain a set of 20 weights (column 5). These weights are multiplied by a set of sampled standardized program values (column 4) to generate a program rating (sum of column 6). This process is repeated another 499 times, generating 500 ratings for each of the 117 economics programs. Each of these 500 ratings for the program is *ranked* by comparing it with the ratings for the other 116 economics programs. The 500 rankings for the program are then ordered from highest to lowest, with the 25th value being the 5th percentile ranking and the 475th being the 95th percentile ranking. These values determine the range of rankings for the program for each measure. Ninety per cent of the 500 randomly generated rankings for the program fall within this range[29,30]. A similar process can be carried out for the dimensional rankings and these results are available in the online spreadsheet in Tables 4a through 6b. The dimensional ranges of rankings for all the economics programs are shown in Appendix B.

The importance of correcting for collinearity [31] is evident from the correlation matrix and is shown in Table 5-4. Citations per publication, for example, have a correlation of .7 with awards, and .5 with GRE-Q, with average Ph.D.'s and with percent completing within six years. This interdependence is corrected for by the principal components adjustment described in Appendix A.

[29] Use of the ninety percent range means that we "throw away" ten percent of the possible rankings for the program. The tails of the distribution can be very long, however, and the ninety percent range is useful in making meaningful comparisons, while illustrating the point that any point estimate of a ranking is inexact.

[30] We do not show the 117 x 500 matrix of all the ordered ratings for all the economics programs, although it will be available on request to researchers. The ranking is obtained from that table.

[31] That is, high degrees of correlation among some of the independent variables.

Table 5-4 Correlations Matrix- Economics

	R Rankings: 5th Percentile	*R Rankings: 95th Percentile*	*S Rankings: 5th Percentile*	*S Rankings: 95th Percentile*	*Research Activity: 5th Percentile*	*Research Activity: 95th Percentile*	*Student Support & Outcomes: 5th Percentile*	*Student Support & Outcomes: 95th Percentile*	*Diversity: 5th Percentile*	*Diversity: 95th Percentile*
R Rankings: 5th Percentile	1.00									
R Rankings: 95th Percentile	0.95	1.00								
S Rankings: 5th Percentile	0.80	0.79	1.00							
S Rankings: 95th Percentile	0.84	0.86	0.96	1.00						
Research Activity: 5th Percentile	0.77	0.77	0.94	0.92	1.00					
Research Activity: 95th Percentile	0.82	0.84	0.92	0.96	0.97	1.00				
Student Support & Outcomes: 5th Percentile	0.24	0.27	0.50	0.48	0.29	0.31	1.00			
Student Support & Outcomes: 95th Percentile	0.24	0.28	0.49	0.49	0.32	0.33	0.92	1.00		
Diversity: 5th Percentile	-0.09	-0.11	-0.16	-0.18	-0.19	-0.21	-0.19	-0.25	1.00	
Diversity: 95th Percentile	-0.14	-0.17	-0.19	-0.20	-0.22	-0.24	-0.17	-0.23	0.93	1.00
Average Number of Publications (2000-2006) per Allocated Faculty, 2006	-0.52	-0.50	-0.75	-0.73	-0.81	-0.79	-0.29	-0.31	0.22	0.26
Average Citations per Publication	-0.77	-0.82	-0.72	-0.80	-0.73	-0.81	-0.26	-0.33	0.15	0.15
Percent of Faculty with Grants, 2006	-0.36	-0.41	-0.47	-0.48	-0.54	-0.53	-0.08	-0.09	0.13	0.15
Awards per Allocated Faculty Member, 2006	-0.53	-0.61	-0.56	-0.65	-0.54	-0.62	-0.34	-0.45	0.23	0.23
Percent of First Year Students with Full Financial Support, Fall 2005	-0.15	-0.15	-0.33	-0.34	-0.16	-0.21	-0.74	-0.64	0.02	-0.01
Average Completion Ratio, 6 Years or Less	-0.39	-0.42	-0.51	-0.50	-0.43	-0.43	-0.54	-0.62	0.30	0.32
Median Time to Degree (Full- and Part-Time Graduates), 2006	0.06	0.07	0.21	0.15	0.17	0.13	0.34	0.38	0.01	0.02
Percent with Academic Plans	-0.12	-0.16	-0.31	-0.29	-0.15	-0.15	-0.57	-0.59	0.26	0.21
Collects Data About Post-Graduation Employment (1=Yes; 0=No)	0.03	-0.02	-0.02	-0.02	-0.03	-0.04	-0.25	-0.25	0.11	0.15
Non-Asian Minority Faculty as a Percent of Total Core and New Faculty, 2006	-0.03	0.00	-0.05	-0.05	-0.03	-0.04	-0.12	-0.09	-0.15	-0.26
Female Faculty as a Percent of Total Core and New Faculty, 2006	0.22	0.23	0.23	0.22	0.16	0.15	0.27	0.25	-0.32	-0.36
Non-Asian Minority Students as a Percent of Total Students, Fall 2005	0.10	0.11	0.09	0.10	0.12	0.12	-0.03	0.00	-0.30	-0.46
Female Students as a Percent of Total Students, Fall 2005	0.39	0.40	0.29	0.33	0.30	0.33	0.27	0.30	-0.38	-0.41
International Students as a Percent of Total Students, Fall 2005	-0.10	-0.07	0.01	0.01	0.05	0.06	0.09	0.11	-0.83	-0.72
Average Number of Ph.D.s Graduated, 2002-2006	-0.69	-0.76	-0.48	-0.57	-0.50	-0.57	-0.02	-0.04	0.05	0.14
Percent of Interdisciplinary Faculty, 2006	0.03	0.01	-0.11	-0.09	-0.02	-0.01	-0.17	-0.10	0.13	0.14
Average GRE Scores, 2004-2006	-0.79	-0.72	-0.66	-0.66	-0.54	-0.56	-0.26	-0.24	-0.04	0.03
Percent of Students with External Fellowships, 2005	-0.16	-0.20	-0.25	-0.28	-0.21	-0.23	-0.21	-0.23	0.11	0.06
Is Student Work Space Provided? (1=Yes; 0=No)	0.01	0.00	-0.27	-0.18	-0.21	-0.14	-0.29	-0.29	0.12	0.12
Is Health Insurance Provided? (1=Yes; 0=No)	-0.24	-0.27	-0.26	-0.26	-0.23	-0.26	-0.17	-0.14	0.15	0.14
Number of Student Activities (Max=18)	-0.19	-0.24	-0.10	-0.10	-0.04	-0.07	0.01	0.05	-0.05	0.01

	Average Number of Publications (2000-2006) per Allocated Faculty, 2006	*Average Citations per Publication*	*Percent of Faculty with Grants, 2006*	*Awards per Allocated Faculty Member, 2006*	*Percent of First Year Students with Full Financial Support, Fall 2005*	*Average Completion Ratio, 6 Years or Less*	*Median Time to Degree (Full- and Part-Time Graduates), 2006*	*Percent with Academic Plans*	*Collects Data About Post-Graduation Employment (1=Yes; 0=No)*	*Non-Asian Minority Faculty as a Percent of Total Core and New Faculty, 2006*
R Rankings: 5th Percentile										
R Rankings: 95th Percentile										
S Rankings: 5th Percentile										
S Rankings: 95th Percentile										
Research Activity: 5th Percentile										
Research Activity: 95th Percentile										
Student Support & Outcomes: 5th Percentile										
Student Support & Outcomes: 95th Percentile										
Diversity: 5th Percentile										
Diversity: 95th Percentile										
Average Number of Publications (2000-2006) per Allocated Faculty, 2006	1.00									
Average Citations per Publication	0.43	1.00								
Percent of Faculty with Grants, 2006	0.35	0.37	1.00							
Awards per Allocated Faculty Member, 2006	0.41	0.67	0.28	1.00						
Percent of First Year Students with Full Financial Support, Fall 2005	0.19	0.15	0.06	0.19	1.00					
Average Completion Ratio, 6 Years or Less	0.31	0.49	0.09	0.52	0.24	1.00				
Median Time to Degree (Full- and Part-Time Graduates), 2006	-0.19	-0.09	-0.06	-0.05	-0.05	-0.33	1.00			
Percent with Academic Plans	0.15	0.14	0.03	0.29	0.11	0.22	-0.03	1.00		
Collects Data About Post-Graduation Employment (1=Yes; 0=No)	0.05	0.06	0.01	0.03	0.02	0.01	-0.07	0.07	1.00	
Non-Asian Minority Faculty as a Percent of Total Core and New Faculty, 2006	0.05	-0.02	0.12	-0.04	0.17	-0.17	-0.12	0.13	-0.08	1.00
Female Faculty as a Percent of Total Core and New Faculty, 2006	-0.28	-0.04	0.01	-0.21	-0.17	-0.13	-0.03	-0.27	0.00	0.10
Non-Asian Minority Students as a Percent of Total Students, Fall 2005	-0.08	-0.12	-0.18	0.01	-0.01	-0.13	-0.08	0.11	-0.08	0.41
Female Students as a Percent of Total Students, Fall 2005	-0.24	-0.28	-0.05	-0.30	-0.04	-0.36	0.01	-0.25	-0.23	0.05
International Students as a Percent of Total Students, Fall 2005	-0.09	-0.01	-0.10	-0.08	0.02	-0.13	-0.01	-0.18	-0.05	-0.10
Average Number of Ph.D.s Graduated, 2002-2006	0.33	0.62	0.38	0.56	-0.07	0.28	0.10	0.04	-0.03	-0.04
Percent of Interdisciplinary Faculty, 2006	-0.09	0.03	0.04	0.19	0.08	0.10	0.06	0.14	0.03	-0.02
Average GRE Scores, 2004-2006	0.37	0.53	0.16	0.38	0.23	0.34	0.04	0.11	-0.05	-0.19
Percent of Students with External Fellowships, 2005	0.13	0.37	0.15	0.43	0.12	0.28	-0.03	0.12	0.06	-0.14
Is Student Work Space Provided? (1=Yes; 0=No)	0.18	0.08	0.12	-0.07	0.23	0.26	-0.26	0.09	0.01	0.04
Is Health Insurance Provided? (1=Yes; 0=No)	0.19	0.21	0.13	0.12	0.17	0.06	0.01	0.10	0.10	0.10
Number of Student Activities (Max=18)	-0.08	0.10	0.12	0.00	-0.05	-0.02	0.09	0.10	-0.01	0.15

	Female Faculty as a Percent of Total Core and New Faculty, 2006	*Non-Asian Minority Students as a Percent of Total Students, Fall 2005*	*Female Students as a Percent of Total Students, Fall 2005*	*International Students as a Percent of Total Students, Fall 2005*	*Average Number of Ph.D.s Graduated, 2002-2006*	*Percent of Interdisciplinary Faculty, 2006*	*Average GRE Scores, 2004-2006*	*Percent of Students with External Fellowships, 2005*	*Is Student Work Space Provided? (1=Yes; 0=No)*	*Is Health Insurance Provided? (1=Yes; 0=No)*	*Number of Student Activities (Max=18)*
R Rankings: 5th Percentile											
R Rankings: 95th Percentile											
S Rankings: 5th Percentile											
S Rankings: 95th Percentile											
Research Activity: 5th Percentile											
Research Activity: 95th Percentile											
Student Support & Outcomes: 5th Percentile											
Student Support & Outcomes: 95th Percentile											
Diversity: 5th Percentile											
Diversity: 95th Percentile											
Average Number of Publications (2000-2006) per Allocated Faculty, 2006											
Average Citations per Publication											
Percent of Faculty with Grants, 2006											
Awards per Allocated Faculty Member, 2006											
Percent of First Year Students with Full Financial Support, Fall 2005											
Average Completion Ratio, 6 Years or Less											
Median Time to Degree (Full- and Part-Time Graduates), 2006											
Percent with Academic Plans											
Collects Data About Post-Graduation Employment (1=Yes; 0=No)											
Non-Asian Minority Faculty as a Percent of Total Core and New Faculty, 2006											
Female Faculty as a Percent of Total Core and New Faculty, 2006	1.00										
Non-Asian Minority Students as a Percent of Total Students, Fall 2005	0.07	1.00									
Female Students as a Percent of Total Students, Fall 2005	0.18	0.05	1.00								
International Students as a Percent of Total Students, Fall 2005	0.06	-0.04	0.05	1.00							
Average Number of Ph.D.s Graduated, 2002-2006	-0.08	-0.10	-0.27	0.06	1.00						
Percent of Interdisciplinary Faculty, 2006	-0.08	0.01	0.00	-0.14	0.11	1.00					
Average GRE Scores, 2004-2006	-0.26	-0.21	-0.15	0.24	0.44	0.10	1.00				
Percent of Students with External Fellowships, 2005	-0.19	0.04	-0.08	-0.03	0.20	0.05	0.14	1.00			
Is Student Work Space Provided? (1=Yes; 0=No)	-0.13	-0.04	-0.04	-0.08	-0.15	-0.02	0.12	0.06	1.00		
Is Health Insurance Provided? (1=Yes; 0=No)	-0.12	-0.17	-0.11	-0.01	0.16	0.03	0.22	-0.21	-0.05	1.00	
Number of Student Activities max=18	-0.04	0.02	-0.11	0.05	0.16	0.12	0.13	-0.08	-0.05	0.19	1.00

The overall ranges of rankings should be looked at in the context of the dimensional measures for economics shown in Appendix B ("IHE University" is program number 62 in the table). Typically, programs that score well on the overall rankings will also do well on the research activity ranking, because the two have a number of highly weighted components in common. It is also worthwhile to look at the student support and outcomes ranking and the diversity ranking, because these may be of importance to students in selecting a program. The economics program's overall measures place it between the 53rd and 86th of the 117 programs for the R-measure and between 27th and 54th for the S-measure . Looking at the dimensional rankings, its research activity is highly ranked—between the 16th and 39th—primarily because of a relatively high rate of publications per allocated faculty member. It does less well in terms of student support and outcomes, where it ranks between the 51st and 95th. Nor does it perform especially well on the diversity dimensional measure—its rank is between the 53rd and 86th. The dimensional measures, then, indicate the specific areas in which programs are performing well or poorly, as separate from the overall ranges of rankings.

The example is intended to explain to the reader how ratings are calculated, and how a range of rankings is constructed. Users are able to click through for any ranked program and gain access to tables similar to the tables above, showing the program data, the range of coefficients for each variable, and the calculation of the 5th and 95th percentile ranking for all the measures for each program. The user should be aware, however, that he or she cannot duplicate all 500 samples of coefficients. Because the ratings depend on program data and weights, both of which have uncertainties associated with them, the ranking resulting from a simulation can only be approximate. The committee would advise that the calculations are more useful in an illustrative sense. That is, for the numerous programs that fall in the middle range of rankings, it doesn't make sense to focus on an exact range. It does make sense to identify the variables that are important to the ranking of each program and, where possible, improve them[32].

[32] An example would be working to shorten time to degree or increasing the percentage of matriculants who receive their degree within 6 years (or 8 years for the humanities).

APPENDIX A

A Technical Discussion of the Process of Rating and Ranking Programs in a Field.

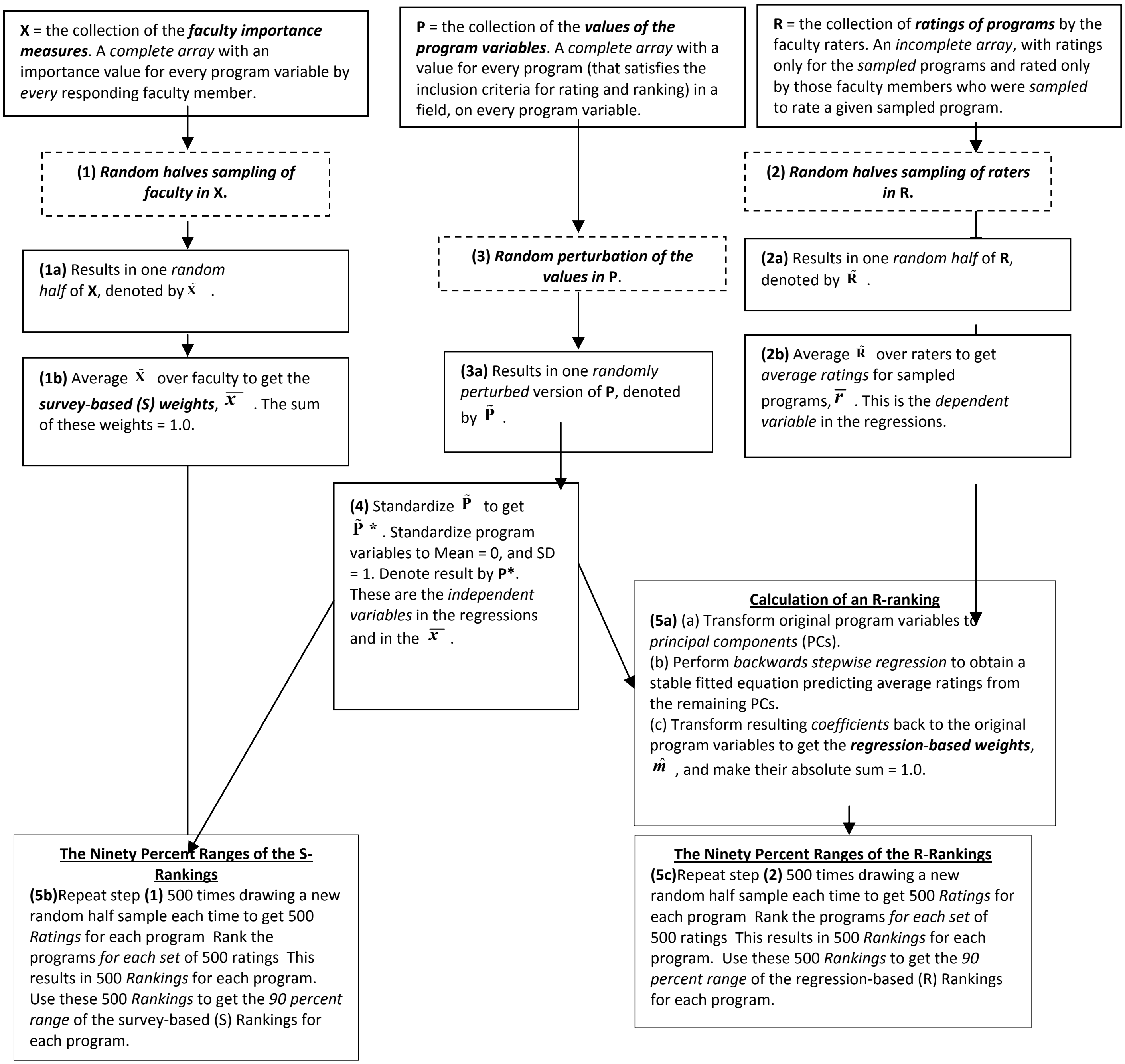
This appendix explains in detail how the various parts of the rating and ranking process for graduate programs fit together and how the process is carried out. Figure A-1 provides a graphical overview of the entire process and forms the basis for this appendix. The appendix addresses each of the boxes in Figure A-1 separately, starting at the top and generally working downward and to the right. The topics in this appendix include:

- a summary of the sources of data used in the rating and ranking process,
- the survey (S)- based weights, the regression (R)-based weights, and the details of the calculations of the endpoints of the 90 percent ranges
- the simulation of the uncertainty in the weights by random-halves sampling,
- the simulation of the uncertainty in the values of the program variables,
- the combination of the simulated weights for the significant program variables with the simulated standardized values of the program variables to obtain simulated rankings, and
- the resulting 90 percent ranges of rankings that are the primary rating and ranking quantities that we report.
- a description of an alternative ranking methodology that combines measures of interest to the user.

THE METHOD FOR CALCULATING THE R AND S RANKINGS

Figure A-1 A graphical summary of the NRC's approach to rating and thereby ranking graduate programs.

The three sets of data: X, P and R.



The Three Data Sets

The empirical basis of the NRC ratings and rankings are the three data sets indicated in the three unlabeled boxes at the top of Figure A-1. The first, denoted by **X**, is the collection of faculty *importance measures* that were derived from data that were collected in the faculty questionnaire. The data in **X** are used to derive the *direct or survey-based (S) weights* discussed more extensively below. The second data set, denoted by **P**, is the collection of the values of the 20 *program variables* that were collected from various sources for each program. The data in **P** are used in the final ratings and rankings of the programs and are discussed in greater detail below. The third, denoted by **R**, is the collection of *ratings of programs by faculty raters*. These ratings were made separately from the faculty questionnaire and involved only a sample of programs from each field and only a sample of faculty raters from that field. This sample of faculty ratings plays a crucial role in the derivation of the *regression-based weights*, discussed more extensively below. More details about these three data sets are also available in Section 2 of this report.

Box (1b): The Direct Weights From the Faculty Questionnaire[1]

Let us turn first to the *survey (S) or direct weights* in box (1b) in Figure A-1, leaving boxes (1) and (1a) to the later discussion of how the uncertainty in these data was simulated.

The faculty questionnaire asks each graduate-program faculty respondent to indicate how important each of 21 characteristics is to the quality of a program in his or her field of study.[2] This information is then used to derive the *survey (S) or direct weights* for each surveyed faculty member, as described below.

The original 21 program characteristics listed on the faculty questionnaire are shown in Table A-1, and they were divided into three categories—faculty, student, and program characteristics. Of the original 21, there are 20 for which adequate data were deemed to be available to use in the rating process, and these 20 data values for each program became the 20 *program variables* used in this study to which we repeatedly refer.

Faculty respondents were first asked to indicate up to four characteristics in each category that they thought were "most important" to program quality. Each characteristic that was listed received an initial score of 1 for that faculty respondent. These preferences were then narrowed by asking the faculty members to further identify a maximum of *two* characteristics in each category that they thought were the most important. Each of these selected characteristics received an additional point, resulting in a score of 2. Given this approach, at most, 12 of the program characteristics can have a non-zero value for any given faculty member; and of these 12, 6, at most, will have a score of 2, and the rest will have a score of 1. At least 8 program characteristics will have a score of 0 for each faculty respondent, more than 8 would be zero if the respondent selected less than 4 as the "important" or 2 as the "most important" characteristics. A final question asked faculty respondents to indicate the *relative importance* of

[1] The importance of program attributes to program quality is surveyed in Section G of the faculty questionnaire.
[2] The number of student publications and presentations was not used because consistent data on it were unavailable. The direct or survey-based and regression-based weights were calculated without it.

each of the three categories by assigning them values that summed to 100 over the three categories.[3] For each faculty respondent, his or her *importance measure* for each program characteristic was calculated as the product of the score that it received times the relative importance value assigned to its category. Finally, the 20 importance measures for each faculty respondent were transformed by dividing each one by the sum of his or her importance measures across the 20 program variables.

Table A-1. The 21 Program Characteristics Listed in the Faculty Questionnaire

Faculty characteristics

i. Number of publications per faculty member

ii. Number of citations per publication (for non-humanities fields)

iii. Percent of faculty holding grants

iv. Involvement in interdisciplinary work

v. Racial/ethnic diversity of program faculty

vi. Gender diversity of program faculty

vii. Reception by peers of a faculty member's work as measured by honors and awards

Student characteristics

i. Median GRE scores of entering students

ii. Percentage of students receiving full financial support

iii. Percentage of students with external funding

iv. Number of student publications and presentations (not used)

v. Racial/ethnic diversity of the student population

vi. Gender diversity of the student population

vii. A high percentage of international students

Program characteristics

i. Average number of Ph.D.'s granted in last five years

ii. Percentage of entering students who complete a doctoral degree in a given time (6 years for non-humanities, 8 years for humanities).

iii. Time to degree

iv. Placement of students after graduation (percent in either positions or postdoctoral fellowships in academia)

v. Percentage of students with individual work space

vi. Percentage of health insurance premiums covered by institution or program

vii. Number of student support activities provided by the institution or program

We will use the following notation consistently: i for a *faculty respondent*, j for a *program* in a field, and k for one of the 20 *program variables*. Thus, x_{ik} denotes the measure of importance placed on program variable k by faculty respondent i. The values, x_{ik}, are non-negative and, over k, sum to 1.0 for each faculty respondent i. The *importance measure vector* for faculty respondent i is the collection of these 20 values,

$$\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{i20}). \tag{1}$$

The entries in these ***x***-vectors are non-negative and sum to 1.00. Denote the vector of *average importance weights*, averaged across the entire set of faculty respondents in a field, by

$$\bar{\boldsymbol{x}} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{20}). \tag{2}$$

The mean value, $\bar{x}_k$, is the average weight of the importance given to the k^{th} program variable by all the surveyed faculty respondents in the field. The averages, $\{\bar{x}_k\}$, are the *direct or survey-based weights* of the faculty respondents because they directly give the average relative importance of each program variable, as indicated by the faculty questionnaire responses in the field of study. Thus, the final 20 importance measures of the program characteristics for each faculty respondent are non-negative and sum to 1.0.

Boxes (2b), (4): The Regression-Based Weights

We next consider the processes in boxes (2b) and (4) in Figure A-1 that lead to the *regression-based weights*. Again, we leave boxes (2) and (2a) to our later discussion of how we simulated the uncertainty in these data.

The regression-based weights represent our attempt to ascertain how much weight is implicitly given to each program variable by faculty members when they rate programs by using their own *perceived* quality of the programs they are rating. We used linear regression to predict average faculty ratings from the 20 program variables and interpreted the resulting regression coefficients as indicating the *implicit importance* of each program variable for faculty ratings. This is different from the survey or direct weights that were just described. We have broken down the process of obtaining the regression-based weights into the three parts indicated by boxes (2b) and (4) which we now discuss in turn.

Box (2b): The average ratings for the sampled programs

The ratings data in R of Figure A-1 are the ratings given by the sampled faculty members to the sample of programs that they were requested to rate. A randomly selected faculty member, i, rates a randomly selected program, j, on a scale of 1 to 6 in terms of his or her *perception* of its quality. Denote this rating by r_{ij}. The matrix sampling plan used was designed so that a sample of up to 50 of the programs in a field was rated by a sample of the graduate faculty members in that same field. Each rater rated about 15 programs, and none rated his or her own program. On average, each rated program was rated by about 44 faculty raters. The rater sample was stratified to ensure proportionality by geographic region, program size (measured by number of faculty), and academic rank. The program sample was stratified to ensure proportionality by geographic region and program size.

R is the array of all the values of r_{ij}. Note that R is an *incomplete array* because many faculty members who responded to the questionnaire did *not* rate programs and many programs in a field were *not* rated, except for the small fields. Box (2b) indicates that we compute the average of these ratings for program *j*, and denote this average rating by $\bar{r}_j$. Because each program's average rating is determined by a different random sample of graduate faculty raters, it is highly unlikely that any two programs will be evaluated by exactly the same set of raters. Denote the *vector* of the average ratings for the sampled programs in a field by $\bar{\boldsymbol{r}}$.

The values of the average ratings in $\bar{\boldsymbol{r}}$ are the *dependent variable* in the regression analyses used to form the regression-based weights.

Box (4): The program variables and standardizing

Denote the value of program variable *k* for program *j* by p_{jk}, and define the vector of all program variables for program *j* by

$$\boldsymbol{p}_j = (p_{j1}, p_{j2}, \ldots, p_{j20}), \qquad (3)$$

and the array with rows given by $\boldsymbol{p}_j$ by P. A cursory examination of the program characteristics listed in Table A-1 shows that they are on *different scales*. For example, the number of publications per faculty member (numbers in the fives and tens), the median GRE scores of entering students (numbers in the hundreds), and the percentage of entering students who complete a doctoral degree in 10 years or less (fractions) are reported in values that are of very different *orders of magnitude*. If these values are left as they are, the size of any regression coefficient based on them will be influenced by *both* the importance of that program variable for predicting the average ratings (which is what we are interested in), as well as the scale of that variable (which is arbitrary and does not interest us). The program variables with *large values*, such as the median GRE scores, will have very small coefficients to reflect the change in scale in going from GRE scores (in the hundreds) to ratings (in the 1 to 6 range). Conversely, program variables with *small values*, such as proportions, will have larger regression coefficients to reflect the change in scale in going from numbers less than 1 to ratings (in the 1 to 6 range).

To avoid the ambiguity between the influence of the scale and the real predictive importance of a variable, we needed to modify the values of the different program variables so they have *similar scales*. This would ensure that program variables with the same influence on the prediction of faculty ratings would have similar regression-coefficient values. Our solution is the very common one of *standardizing* the p_{jk}-values by subtracting their mean across the programs in a field and dividing by the corresponding standard deviation. This will result in program variables that have the same mean (0.0) and standard deviation (1.0) across the programs in the field. In this way, no program variable will have substantially larger or smaller values than any other program variable across the programs in a field. For the regressions of box (4), the standardization was done only over the programs that were sampled for rating.

We denote the values of the standardized program variables with an asterisk (p_{jk}* and **P***). Two program variables (Student Work Space and Health Insurance) were coded as 1 (present) or -1 (absent). We felt that there was no need for additional standardization of these two program variables and they were not standardized to have mean 0 and variance 1.

The standardized program variables for the sampled and rated programs served as the *predictor or independent variables* in the regressions that lead to the regression-based weights.

Box (5a): The regressions and the regression-based weights

The statistical problem addressed in box (4) is to use $\bar{\boldsymbol{r}}$ and **P*** as the *dependent* and *independent* variables, respectively, in a linear regression, to obtain the vector of regression-based weights, $\hat{\boldsymbol{m}}$, using least squares. It should be noted that only the data in **P*** for the *sampled* programs are used. The data for the non-sampled programs in **P*** are not used in this step of the process.

Two immediate problems arise. These are: (1) the number of observations (i.e., the number of sampled programs in a field) is 50 or less, while the number of independent variables (i.e., the program variables in **P***) is 20, and (2) a number of the program variables are correlated with each other across the programs in a field. This is less than an ideal situation for obtaining *stable* regression coefficients. There are too few observations to hope for stable estimates of the coefficients for 20 variables. The fact that these variables are also correlated does not help matters either. If we had ignored these two problems, least-squares regression methods would have tended to assign coefficients rather arbitrarily to one particular variable or to other variables that are correlated with it, and how this worked out would depend on which programs were included in the sample of rated programs. The resulting unstable regression coefficients would have been unusable for our purposes.

For example, as expected, when we fit a linear model that included all 20 of the program variables, we found that for a number of the variables, the coefficients and their signs did not make intuitive sense. However, we found, as expected, that they made more sense when we used various step-wise selection methods for reducing the number of variables used as predictors. With only 50 cases, we had to expect that we could not use all 20 variables in the prediction equations without adjustments.

After examining a variety of approaches, we settled on using a backwards, step-wise selection method applied to the 20 *principal component* (PC) variables formed from the 20 program variables (rather than using the original 20 program variables). The regression coefficients obtained for the remaining PC variables were then transformed back to scale of the original 20 program variables, with the result that all 20 program variables now had non-zero coefficients, but these coefficients were subject to several linear constraints implied by the deleted PC variables.

The principal component variables are linear combinations of the original 20 program variables that have two properties: (1) they are uncorrelated in the sample, and (2) they can give exactly the same predictions as do the original variables—that is, every prediction equation that is possible with the original variables is also possible to form using the PC variables, using different regression coefficients. The PC variables are usually ordered by their variances from largest to smallest, but this plays no role here. There are as many PC variables as there are original variables—in our case, 20.

If we denote the array of original 20 standardized variables for the sample of rated programs as **P***, then the corresponding array of the 20 PC variables, **C**, is given by the matrix multiplication, **C** = **P*V**, where **V** is the 20 by 20 orthogonal matrix specified by, among other things, the *singular value decomposition* of **P***. After the regression coefficients are estimated using the PC variables, we get back to the coefficients for the original standardized variables in **P*** by transforming the vector of regression coefficients by the transformation, **V**.

Our step-wise use of the PC variables proceeded as follows. We begin with a least-squares prediction equation, predicting $\bar{\boldsymbol{r}}$ from **C**, that includes all of the PC variables. Then a series of analyses is performed, with one PC variable at a time being left out of the prediction equation; the PC variable that has the least impact on the fit of the predicted ratings (as measured by its t-statistic) is removed. This process is repeated, removing one PC variable each time, until the remaining PC variables each add statistically significant improvements to the fit of the predictions of the ratings (at the 0.05 level). The result is a set of regression coefficients, the *PC coefficients*, $\hat{\gamma}$, which predict the sample of program ratings from a subset of the PC variables, i.e.,

$$\hat{\bar{\boldsymbol{r}}} = \mathbf{C}\hat{\tilde{\boldsymbol{\gamma}}} \tag{4}$$

In Equation 5, the caret denotes estimation. Moreover, for the PC variables that have been eliminated during the backwards selection process, the corresponding PC-coefficients, $\hat{\gamma}_{\boldsymbol{k}}$, are zero. These zeros mean that we are setting the *coefficients* of certain *linear combinations of the original variables* to zero rather than setting the coefficients for some of the original program variables to zero. This was regarded as a virtue, because we did not *necessarily* eliminate any of the original program variables from the prediction equation used to find the regression-based weights. By proceeding this way, we are not forced to give a zero weight to one of two collinear variables in the step-wise procedure. Instead, both collinear variables will typically load onto the same principal components and get some weight when the matrix **V** is applied to the PC coefficients to obtain the coefficients for the original program variables, i.e.,

$$\hat{\boldsymbol{m}} = \mathbf{V}\hat{\boldsymbol{\gamma}}\,. \tag{5}$$

In the same way, the matrix of estimated variances and covariances of $\hat{\boldsymbol{\gamma}}$, obtained from the least-squares output, may be transformed to the corresponding matrix for $\hat{\boldsymbol{m}}$.[4]

The regression coefficient for the k^{th} program variable, denoted by $\hat{m}_k$, is the *regression-based weight* for program characteristic *k* as a predictor of the average ratings of the programs by the faculty raters, and $\hat{\boldsymbol{m}} = (\hat{m}_1, \hat{m}_2, ..., \hat{m}_{20})$.

The predicted perceived quality rating for a sampled program can be expected to *differ* somewhat from the actual average rating for that program. For example, for the two fields

[4] If the weights from the R and S measures were to be combined, the variances from this matrix would be used later [in box (6) of the computation of combined weights] in the computation of the "optimal fraction" for combining the survey-based and regression-based weights.

studied in *Assessing Research Doctorate Programs: A Methodology Study*, the root-mean-square deviation between the predictions and the average ratings was 0.42 on a 1-to-6 rating scale for both mathematics and English. In addition, the (adjusted) R^2 of the regressions of average ratings on measured program characteristics was 0.82 for mathematics and 0.80 for English. These values indicate that the predictions account for about 80 percent of the variability in average ratings. We regarded this as satisfactory levels of agreement between predicted and actual to use these methods in this study.

These results show that the *predicted* perceived quality ratings agree fairly well with the *actual* ratings. However, these results do not indicate how well a prediction equation that was based on a *sample of programs* will reproduce the predictions of the equation for the *whole population of programs* in a field. The data for mathematics, reported in *Assessing Research Doctorate Programs: A Methodology Study*, indicate that using 49 programs did a reasonably good job of reproducing the predictions based on the whole field of 147 mathematics programs.[5] Thus, we decided that in developing the regression-based ratings, we would use a sample of 50 programs from a field if it had more than 50 programs and use almost all of the programs in fields with 50 or fewer programs. When there were fewer than 30 programs in a field, it was combined with a larger discipline with similar direct weights for the purposes of estimating the regression-based weights.[6] In two cases, computer engineering and engineering science and materials, there were fewer than 25 programs, and these fields were not ranked, although data are reported for all 20 characteristics.[7]

There is one final alteration in the values of $\hat{\boldsymbol{m}}$ that needs to be mentioned. The survey-based or direct weights, $\{\bar{x}_k\}$, have absolute values that sum to 1.0. This is not necessarily true of the regression coefficients, $\{\hat{m}_k\}$. The scale of m_k depends on both the scale of p_{jk} and the scale of the average ratings, $\{\bar{r}_j\}$.We decided, because initially our intent was to *combine* these two sources of the importance of the various program variables, that they needed to be on similar

---

[5] See Appendix G of *Assessing Research Doctorate Programs: A Methodology Study,* National Research Council (2003)
[6] The fields for which this was done were:

Small Field	Surrogate Field
Aerospace engineering	Mechanical engineering
Agricultural economics	Economics
American studies	English literature
Astrophysics and astronomy	Physics
Entomology	Plant science
Forestry	Plant science
Food science	Plant science
Theatre and performance	English literature

[7]Ranges of rankings are not provided for three fields that were in the original taxonomy: 1)Languages, Societies, and Cultures, for which the sub-fields were too diverse to it as a coherent field; and 2)Engineering Science and Materials and 3) Computer Engineering, which fell below the minimum of 25 programs to permit the calculation of rankings for a field. The committee had not anticipated this when it developed the taxonomy, or the fields would not have been included as a separate field.

scales. We decided to force them *both* to sum to 1.0 in absolute value[8]. This allows the direct and regression-based weights to have negative values where they arise, typically in the regression-based weights, without requiring anything complicated to deal with this. Using the sum of absolute values allows the sign of the regression-based weights to be determined by the data rather than by an a priori hypothesis. Thus, we divided each regression coefficient, $\hat{m}_k$, by the sum of the absolute values of all the regression coefficients. In this way, both the direct and regression-based weights are fractional values, mostly positive but some negative, whose absolute sums equal 1.0[9].

Boxes (1), (1a), (2) and (2a): Simulating the Uncertainty in the Direct and Regression-Based Weights

The survey-based (S) or direct weight vector, $\bar{\boldsymbol{x}}$, is subject to uncertainty; that is, a different set of respondent faculty would have led to different values in $\bar{\boldsymbol{x}}$. Disagreement among the graduate faculty on the relative importance of the 20 program variables is the source of the uncertainty of the direct or survey-based weights. The average ratings of the sampled faculty in $\bar{\boldsymbol{r}}$ are also subject to uncertainty; a different sample of raters or programs would have produced different values in $\bar{\boldsymbol{r}}$. One way to reflect this uncertainty is to use the sampling distributions of $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{r}}$. There are various ways that these sampling distributions may be realized. We chose an empirical approach that made no assumptions about the shapes of the various distributions involved, but this allowed us to use computer-intensive methods to let the sampling variability of both $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{r}}$ influence the final ratings and rankings. We examined two empirical approaches, Efron's *bootstrap* and a *random-halves* (RH) procedure suggested by the committee chairman. We found that both gave very similar final results in terms of the final ranges of rankings and ratings. The bootstrap requires taking a sample of *N* with replacement from the relevant empirical distribution. The RH procedure requires taking a sample of *N*/2 without replacement from the same empirical distribution. We chose to use the RH procedure because it cut the sampling computations in half, is fairly easy to explain, and as far as we could tell, gave essentially the same results as the bootstrap for ranking and rating.

Boxes (1) and (2): The random halves procedure

The RH procedure for both $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{r}}$ are nearly the same, and with the same justifications. **X** is a complete array whose rows denote the *N* faculty respondents, while **R** is an incomplete array whose rows denote the *n* sampled faculty raters for a field. In the case of **X**, the RH procedure requires a random sample of size *N*/2 of the *faculty respondents*. In the case of **R**, the RH

[8] We use the absolute value here because, for time to degree, a higher value should receive a negative weight. Note that normalization has no effect on relative rankings, since it is simply a linear transformation.

[9] The estimated standard deviations of the {$\hat{m}_k$}, obtained in standard ways from the regression output, were also divided by this sum to make them the correct size for use in the process of combining the direct and regression-based weights, discussed below.

procedure requires a random sample of size $n/2$ of the *faculty raters*. Repeated draws from these random half samples are then used to simulate the uncertainty in $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{r}}$, respectively.

Alert readers may worry that these half samples will exhibit *too much* variability in the resulting averages; after all, a half sample has only half the number of cases as a full sample—and the bootstrap always takes a full sample of N or n. The explanation of why a half sample without replacement has essentially the same variability as a full sample with replacement is most easily seen by considering the variance of the mean of a sample without replacement from a finite population. It is well known from sampling theory that the variance of the mean from a sample of size $N/2$, from a population of size N is, essentially,

$$\mathrm{Var}(\bar{x}_k) = \frac{\sigma^2_{x_k}}{\left(\frac{N}{2}\right)}\left(1 - \frac{N}{2}/N\right) = \frac{\sigma^2_{x_k}}{N}. \tag{6}$$

That is, because of the "finite sampling correction," the variance from a random half sample without replacement is exactly the same as the variance of a random sample of twice the size with replacement (there is a small "N versus $N - 1$" effect that Formula 11 ignores). This is why the bootstrap and the RH methods give such similar results in our application to the uncertainty of the direct weights. There are other reasons to expect the RH method to produce a useful simulation of the uncertainty of averages.[10]

The same reasoning applies to the RH sampling of the faculty raters in **R** to simulate the uncertainty in the average ratings, $\bar{\boldsymbol{r}}$, used to obtain the regression-based weights. The procedure was to sample a random half of all raters for programs in a field and compute the average rating for each program from that half sample.

The regression-based weights are subject to uncertainty from *two* sources. The first is the uncertainty arising from sampling the faculty raters and, as indicated above, the RH sampling directly addresses this source. The second is from using average ratings from a sample of programs rather than all the programs to develop the regression equation from which the regression-based weights are derived. In the discussion of box (4), above, we gave our reasoning for believing the sample of 50 programs is adequate, and how we pool the data from other related fields when the number of programs in a field is smaller than 50. In addition, while the use of ratings for a sample of programs has the practical value of reducing the workload of the faculty raters, our *implicit* use of the predicted average ratings, $\{M_j\}$, from Equation 5 above, rather than actual average ratings, $\{\bar{r}_j\}$, also reduces some of the uncertainty due to the sampling of the programs to be rated. For these two reasons, we believe that this second source of uncertainty is not as important as that simulated by the RH procedure for the uncertainty in the average ratings, and consequently, for the regression-based weights, $\hat{\boldsymbol{m}}$.

We always drew the RH samples 500 times, and those for $\bar{\boldsymbol{x}}$ were statistically independent of those for $\bar{\boldsymbol{r}}$. This gives us 500 replications of the direct or survey-based weights and 500 replications of the regression-based weights.

---

[10] The random-halves procedure has a place in the statistical literature, but with other names. It is an example of the "deleted-d" jackknife as described in Efron and Tibshirani, (1993) *An Introduction to the Bootstrap.* New York: Chapman and Hall. p. 149, with d = n/2. It is described by Kirk Wolter in a private communication as an example of the "balanced repeated replication" or "balanced half samples," and described in Wolter, K. M. (2007) *Introduction to Variance Estimation*., 2nd ed. New York: Springer-Verlag.

Boxes (3) and (3a): Incorporating Uncertainty into the Program Variables

In addition to the uncertainty in the survey-based (direct) and regression-based weights discussed above, there is also some uncertainty in the values of the program variables themselves. Some of the 20 program variables used to calculate the ratings also vary or have an error associated with their values due to year-to-year fluctuations. Data for five of the variables (publications per faculty, citations per publications, GRE scores, Ph.D. completion, and number of Ph.D.'s) were collected over time, and averages over a number of years were used as the values of these program variables. If a different time period had been used, the values would have been different. To express this type of uncertainty, a *relative error factor,* e_{jk}, was associated with each program variable value, p_{jk}. The relative error factor was calculated by dividing the standard deviation over the series by the square root of the number of observations in the series, and then dividing that number by the value of the variable p_{kj}. For example, the publications per faculty variable is the average number of allocated publications per allocated faculty over 7 years, and a standard error value was calculated for this variable as SD/$\sqrt{7}$. This standard error was then divided by the value of the publications per faculty variable to get the relative error factor for this program variable.

For the other 15 program variables that are used in the ratings, no data on variability were directly obtained during the study, and we assigned a relative error of 0, 0.1 or 0.2 to these variables. The relative error for the variables Student Workspace and Health Insurance were given an error of 0, because they were thought to have little or no temporal fluctuation over the interval considered; and for Percent of Faculty Holding Grants, the error assigned was 0.2, because an examination of data from the *National Science Foundation Survey of Research Expenditure* indicated this to be an appropriate estimate. The remaining 12 program variables were assigned a relative error of 0.1. Each program had its own relative error factor for each program variable, e_{jk}.

Just as we had simulated values from the sampling distributions of $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{r}}$ via RH sampling, we also wanted to reflect the uncertainty in the values of the program variables themselves rather than using the fixed values, $\{p_{kj}\}$, in computing program ratings. We did this in the following way. The value, p_{kj}, was *perturbed* by drawing randomly from the Gaussian distribution, $N(p_{kj}, (e_k p_{kj})^2)$.This distribution has a mean equal to the variable value p_{kj} and a standard deviation equal to the relative error, e_k, times the variable value, p_{kj}. Thus, the entire array $\mathbf{P}$ is randomly perturbed to a new array, $\tilde{\mathbf{P}}$. This perturbing process is repeated 500 times, and each one is standardized to have mean 0.0 and standard deviation 1.0 for each of the 20 program variables to produce 500 standardized arrays, $\tilde{\mathbf{P}}*$.

Boxes (5b) and (5c): The Ninety Percent Ranges of the S and R Rankings

In box (5b) we have already calculated 500 replications of the survey-based weights and in box (5c) we have done the same for the Regression-based weights for the given field [from box (2b)] and from 500 replications of the steps in boxes 5b and 5c we have 500 replications of the standardized perturbed version of **P** that contains the program variable data for all of the programs to be rated in the field.

For either measure, denote the k^{th} replication of R_j by $R_j^{(k)}$. To obtain the k^{th} replication of the *rankings* of the programs, sort the values of $R_j^{(k)}$ over j from high to low and assign the rank of 1 to the program with the highest rating in this set. In case of tied ratings, we use the standard procedure in which the ranks are averaged for the tied cases, and the common rank given to the tied programs is the average of the ranks that would have been given to the tied set of programs. For each of the replications of the ratings, there is a corresponding replication of the rankings of the programs, resulting in 500 replications of the ranking of each program.

Instead of reporting a single ranking of the programs in a field, we report the ninety percent range of the rankings for each program. This is an interval starting with the rank that was at the 5th percentile in the distribution of the 500 replications of the ranks for the given program, and ending at the 95th percentile of this distribution. The interpretation of the ninety percent range is that it is *range that covers the middle ninety percent of the rankings* and reflects the uncertainty in the survey-based (direct) and regression-based weights and in the program data values five percent of a program's rankings in our process are less than this interval and five percent are higher. The interval itself represents what we would expect the typical rankings for that program to be, given the uncertainty in the process and the ratings of the other programs in the field.[11] These ninety percent ranges are reported for the R and S measures, as well as for the three dimensional measures.

AN ALTERNATIVE APPROACH TO CONSTRUCTING RANKINGS: COMBINING THE R AND S MEASURES

The pre-publication version of this revised Methodology Guide appeared in July, 2009 and explained the methodology developed by the committee at that time, that is, one that combined the R-based and S-based measures in a way that will be described below. In July 2009, the committee had estimated ranges of rankings for only a handful of fields and assumed that this method of estimation would be generally satisfactory. In theory it is, but when applied to data for additional fields it became clear that there were some fields for which the range of program rankings based on the S-measure differed considerably from that based on the R-measure.

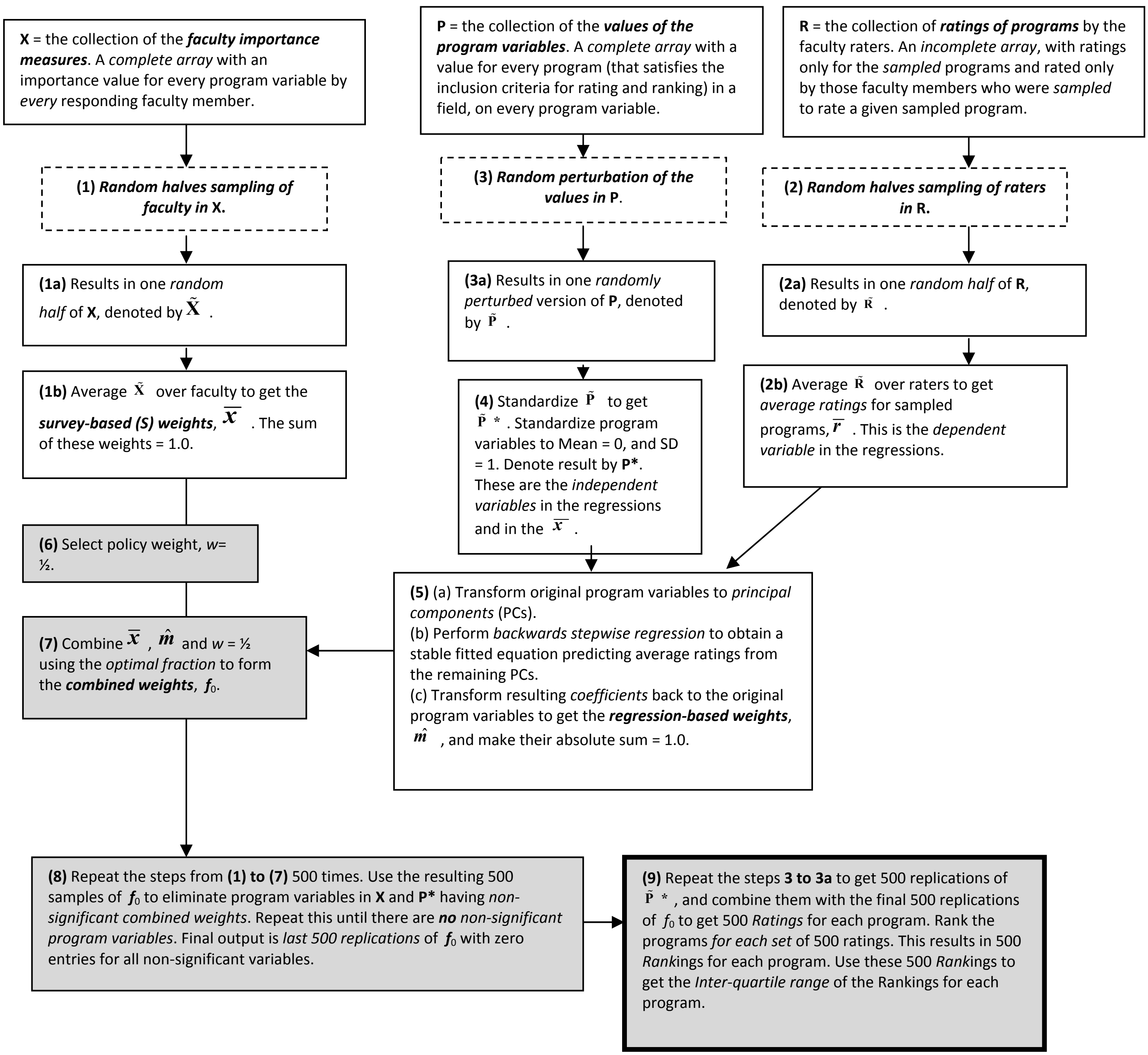
[11]In an earlier draft of this guide, we chose an inter-quartile range, but this choice, rather than some other range (eliminating the top and bottom quintile, for example) is arbitrary. The current approach uses broader ranges which result in greater overlap of ranges, but has the advantage of covering most of the rankings a program might achieve. The point of introducing uncertainty in our calculations is that we do not know the "true" ranking of a program. The purpose of presenting a ninety percent range is to provide a range in which a program's ranking is likely to fall.

Further, the committee came to view any set of ranges of rankings that it might develop as illustrative, that is, any range of rankings depended critically on the characteristics chosen and the weights applied to those characteristics. The R- and S- based ranges of rankings were two examples of data-based ranking schemes, but there are others. In fact, the dimensional measures described in the body of this Guide, are an example[12]. The technical description of further steps that the committee carried out to obtain ranges of rankings using the combined measures are described in this section—beginning with an alternative conceptual diagram.

[12] In most cases, it would not make sense to combine the dimensional measures because they yield differing results for most programs.

Figure A-2 A graphical summary of the alternative method.

The three sets of data: X, P and R.



Note: Shaded boxes indicate steps used in an alternative technique and are omitted from the technique used to generate the current rankings

Boxes (6) and (7): The Combined Weights

To motivate our method of combining of the direct and regression-based weights, we start by describing the direct and regression-based *ratings*. Remembering that the standardized values of the program variables for program *j* are denoted by p_{jk}*, the *direct rating* for program *j*, using the average direct weight vector, $\bar{\boldsymbol{x}}$, is X_j, is given by

$$X_j = \sum_{k=1}^{20} \bar{x}_k p_{jk}^{*}. \qquad (7)$$

The *regression-based rating* for program *j*, using the regression-based weight vector, $\hat{\boldsymbol{m}}$, is M_j, is given by

$$M_j = \sum_{k=1}^{20} \hat{m}_k p_{jk}^{*}. \qquad (8)$$

Note that the regression-based rating is a linear transformation of the predicted ratings used to obtain the regression-based weights, because the constant term of the regression is deleted, and the weights have been scaled by a common value so that their absolute sum is 1.0. The procedure for computing regression-based ratings can be used for any program, sampled or not, in the given field. Simply use M_j *as* defined in Equation 7 above, where $\{p_{jk}^{*}\}$ comes from the data for program *j* and the $\{\hat{m}_k\}$ are the regression-based weights based on the sample of programs and raters.[13]

We combined the direct ratings with the regression-based ratings as follows. Let *w* denote a *policy weight* and form the following *combination* of the direct and regression-based ratings:

$$R_j = wM_j + (1 - w)X_j. \qquad (9)$$

The *policy weight*, *w*, is chosen in box (5) of Figure A-1, and is the amount the regression-based ratings are allowed to influence the combined rating, R_j. When $w = 0$, the regression-based rating has *no* influence on the R_j. When $w = 1$, the R_js are *totally* based upon the regression-based ratings. Any *compromise value* of *w* is somewhere between 0 and 1.

We did not actually form both the direct and regression-based ratings in our work. Instead, we exploited the simple linear form of these given by:

$$R_j = w\sum_{k=1}^{20} \hat{m}_k p_{jk}^{*} + (1 - w)\sum_{k=1}^{20} \bar{x}_k p_{jk}^{*} = \sum_{k=1}^{20} \bar{f}_k p_{jk}^{*} \qquad (10)$$

[13] We have throughout estimated linear regressions. Is this assumption justified? We can only say that, empirically, we tried alternative specifications that included quadratic terms for the most important variables (publications and citations) and did not find an improved fit.

where the combined weight, $\bar{f}_k$, is given by

$$\bar{f}_k = w\hat{m}_k + (1 - w)\bar{x}_k . \qquad (11)$$

The representation of the combined rating given in Equations 9 and 10 is a linear combination of the program variables that uses the *combined weights*, $\{\bar{f}_k\}$ defined in Equation 10. The combined weight $\bar{f}_k$ is applied to the k^{th} standardized program characteristic, p_{jk}* for each k, and then all 20 of these weighted values are summed to obtain the final combined rating for program j.

However, because both $\hat{m}_k$ and $\bar{x}_k$ are subject to uncertainty, we made one additional adjustment to Equation 10 that is described below, following the discussion of how we simulated the uncertainty in both the direct weights and in the average ratings used to form the regression-based weights.

Box (7): Using the optimal fraction to combine the direct and regression-based weights.

In deriving the ranges of ratings that reflect the uncertainty in $\hat{m}_k$ and $\bar{x}_k$, simulated values, m_k, and x_k, are drawn from the sampling distributions of $\hat{m}_k$, and $\bar{x}_k$, respectively, using independent RH samples from the appropriate parts of **R** and **X**. These two simulated values are to be combined to form a simulated value, f_k, for $\bar{f}_k$ in Equation 11. However, the simple weighted average in Equation 11 only reflects the effect of the policy weighting, w, and ignores the fact that both m_k, and x_k are independent random draws from distributions, rather than fixed values. We want to combine m_k, and x_k in such a way as to bring the simulated value, f_k, as close as possible to $\bar{f}_k$ on average, and in a way that will also reflect the policy weight, w, appropriately. This section outlines our approach to choosing the *optimal fraction* to apply to m_k to achieve this. The optimal fraction is the amount of weight applied to m_k that minimizes the mean-square error of f_k, treating $\bar{f}_k$ as a target parameter to be estimated.

First, consider a general weighting, $f_k(u)$, that uses a fraction, u. This weighting has the form

$$f_k(u) = um_k + (1 - u)x_k. \qquad (12)$$

By construction of the RH procedure, the mean of the distribution of m_k is $\hat{m}_k$ (the regression coefficients that are obtained when the data from all n faculty raters are used). Similarly, the mean of the distribution of x_k is $\bar{x}_k$, the mean importance value that is obtained when the data from all N faculty respondents are averaged. We may regard $f_k(u)$ as an estimator of ϕ_k, given by

$$\phi_k = w\hat{m}_k + (1 - w)\bar{x}_k . \qquad (13)$$

The problem then is to find the value of u that will minimize the mean-square error (MSE) of $f_k(u)$ given by

$$\mathrm{MSE}(u) = \mathrm{E}(f_k(u) - \phi_k)^2, \qquad (14)$$

where, in Equation 14, the notation, $\mathrm{E}(f_k(u) - \phi_k)^2$ denotes the *expectation* or *average* taken over the independent RH distributions of $\hat{m}_k$ and $\bar{x}_k$. The MSE is a measure of the combined uncertainty in $f_k(u)$.

The MSE in (14) can be written as

$$\begin{aligned}\mathrm{MSE}(u) &= \mathrm{E}(um_k + (1-u)x_k - w\hat{m}_k - (1-w)\bar{x}_k)^2 \\ &= \mathrm{E}(u(m_k - \hat{m}_k) + (1-u)(x_k - \bar{x}_k) + (u-w)\hat{m}_k + (w-u)\bar{x}_k)^2 \\ &= \mathrm{E}(u(m_k - \hat{m}_k) + (1-u)(x_k - \bar{x}_k) + (u-w)(\hat{m}_k - \bar{x}_k))^2. \qquad (15)\end{aligned}$$

The point of re-expressing Equation 14 as Equation 15 is that now when the squaring is carried out, all of the terms except the squared ones have zero expected values and can be ignored. If we denote the variance of the sampling distribution of $\hat{m}_k$ by $\sigma^2(\hat{m}_k)$ and the variance of $\bar{x}_k$ by $\sigma^2(\bar{x}_k)$, then Equation 15 becomes

$$\mathrm{MSE}(u) = u^2\sigma^2(\hat{m}_k) + (1-u)^2\sigma^2(\bar{x}_k) + (u-w)^2(\hat{m}_k - \bar{x}_k)^2. \qquad (16)$$

It is now a straightforward task to differentiate Equation 16 in u, set the result to zero, and solve for the optimal u-value, u_{0k}, which we call the *optimal fraction*. This calculation results in

$$u_{0k} = \frac{\sigma^2(\bar{x}_k) + w(\hat{m}_k - \bar{x}_k)^2}{\sigma^2(\bar{x}_k) + \sigma^2(\hat{m}_k) + (\hat{m}_k - \bar{x}_k)^2}. \qquad (17)$$

The optimal fraction in Equation 12 has some useful and intuitive properties. It takes on the value w when there is no uncertainty about the direct and regression-based weights. Moreover, w has no influence on the optimal fraction when $\hat{m}_k$ and $\bar{x}_k$ are equal. In that case, the direct weights and regression-based weights on the k^{th} program characteristic are the same, and the optimal fraction combines the two simulated values in a way that is inversely proportional to their variances, so that the value with less variation gets more weight. Note also, that the value in Equation12 is the same for all of the RH simulated values of m_k and x_k.

The two variances in Equation 12, $\sigma^2(\bar{x}_k)$ and $\sigma^2(\hat{m}_k)$, may be found in standard ways. The value of $\sigma^2(\bar{x}_k)$ is given by

$$\sigma^2(\bar{x}_k) = \sigma^2(x_k)/N_F, \qquad (18)$$

where N_F denotes the number of faculty in the field who supply direct weight data, and $\sigma^2(x_k)$ denotes the variance of the individual direct weights given to the k^{th} program variable by these faculty respondents. The value of $\sigma^2(\hat{m}_k)$ is obtained from the regression output that produces $\hat{m}_k$ when the data from all faculty raters in a field are used. Its square root, $\sigma(\hat{m}_k)$is the standard error of the regression coefficient, $\hat{m}_k$. Finally, because we rescaled the $\hat{m}_k$ so that their absolute sum was 1.0, the same divisor must be applied to $\sigma(\hat{m}_k)$to put it on the corresponding scale.

If we now replace the u in Equation 17 with u_{0k} given in Equation 17, we then obtain the combined weight that optimally combines the two simulated values of the weights, m_k, and x_k, into the combined rating, given by

$$R_{0j} = \sum_{k=1}^{20} f_{0k} p^*_{kj} \quad (19)$$

where

$$f_{0k} = u_{0k}m_k + (1 - u_{0k})x_k, \quad (20)$$

and u_{0k} is given by Equation 17. The vector of optimally combined weights is denoted by $\boldsymbol{f}_0$[14].

The values of R_{0j} from Equations 19 and 20 are used as the 500 simulated values of the combined ratings for the purposes of determining the ranking interval ranges for each program that is discussed below.

In performing the RH sampling to mimic the uncertainty in the direct and regression-based weights, it should be emphasized that the random half samples from **X** and **R** were statistically independent. This is our justification for assuming that the random draws, m_k, and x_k, are statistically independent in the calculation of the optimal fraction, u_{0k}.[15]

As a final point, we did realize that the approach to calculating the optimal fraction described above did not take into account any correlation between the direct and regression-based weights for *different* program variables. We did examine a method that did, but it simply produced a matrix version of Equation 12 that reduced to the procedure we used when the program variables were uncorrelated, but was otherwise difficult to implement with the resources available to us.

Box 8: Eliminating Non-Significant Program Variables

After we have obtained the 500 simulated values of the combined weights by applying Equations 17 and 20 to the 500 simulated values for the direct and regression-based weights, we were in a position to examine the distributions of these 500 values of the combined weights for each program variable. The distributions of the combined weights for some of the program variables did not contain zero and were not even near zero. However, other program variables had combined weight distributions that did contain zero. If zero is inside the middle 95 percent of this distribution, we declare the combined weight for that program variable to be *non-significant* for the rating and ranking process (in analogy with the usual way that distributions of parameters are tested for statistical significance). If the combined weight for a program variable is not significantly different from zero, the variable for that coefficient is dropped from further computations. This elimination of program variables required us to recalculate everything above box (8) in Figure A-2. The eliminated program variables are ignored in calculating the direct and

[15] The fact that the raters for each field were a subset of those who answered the faculty questionnaire may confuse some into thinking that our independence assumption may not be justified. This is an unfortunate misunderstanding of the simulation of uncertainty in the rating and ranking process. It is the statistical independence of the two RH sampling processes that matters, nothing else.

regression-based weights for the other variables. New RH samples are drawn, the direct weights are retransformed so that the absolute sum of the remaining direct weights was 1.0, the regressions are re-run using the reduced set of program variables as predictors, and new optimal fractions are computed to combine the direct and regression-based weights. Finally, the 500 simulated combined coefficients are again tested for statistical significance from zero. This process is repeated until a final set of combined weights, each of which is significantly different from zero, is obtained. Only after this testing and retesting process is performed are the final sets of 500 combined coefficients ready for use in the computation of the intervals of rankings that are discussed in box (5) of Figure A-1. The values for the combined weights that correspond to the eliminated variables are set to 0.0 in each of the final 500 simulated values of $\boldsymbol{f}_0$. These 500 vectors of combined weights are used in the production of the ratings that are used to produce the final intervals of rankings for each program, as discussed later.

Empirically, the examination of three fields suggests that this process has two useful effects. First, the middle of the inter-quartile ranges of rankings of programs is changed very little, so that the ranges before eliminating nonsignificant program variables and those after this elimination are centered in nearly the same places[16]. Second, the widths of these inter-quartile ranges are slightly reduced or are unchanged. These are the effects that we would expect from eliminating variables that are having only a noisy effect on the ranking and rating process, and for this reason, we have continued to include box (8) in our rating and ranking process. Nonetheless, the inter-quartile intervals do shift more markedly than the medians, when estimated coefficients are set to zero—largely for those departments near the middle of the rankings. This is because quartile estimates are more variable than median estimates. There are even rare instances in which the intervals calculated both ways do not overlap.

From this point on, the calculation of the ranges of rankings is carried out as described in the section about the R-and S- ranges of rankings.

---

[16] Examination of the effect of this procedure gave correlations between the median rankings with and without the elimination of nonsignificant variables of .99.

APPENDIX B

Dimensional Ranges of Rankings for Programs in Economics and the Overall Ranges

	Research Activity			**Student Support & Outcomes**			**Diversity**			**Overall Measure**			
Program	Research Activity: 5th Percentile	Research Activity: 95th Percentile		Student Support & Outcomes: 5th Percentile	Student Support & Outcomes: 95th Percentile		Diversity: 5th Percentile	Diversity: 95th Percentile		R Rankings: 5th Percentile	R Rankings: 95th Percentile	S Rankings: 5th Percentile	S Rankings: 95th Percentile
1	96	108		60	99		71	100		68	97	100	109
2	55	77		65	99		2	14		62	94	62	81
3	69	91		60	94		68	95		61	93	79	96
4	3	6		30	64		52	91		2	3	3	5
5	102	114		93	108		24	74		74	99	100	109
6	38	64		71	99		96	111		85	113	71	94
7	18	32		19	50		38	77		20	31	20	28
8	18	30		34	73		11	60		28	54	22	30
9	88	108		7	29		65	94		104	113	76	96
10	63	91		101	113		10	46		73	104	94	108
11	11	19		23	56		6	40		14	24	11	18
12	23	36		70	101		74	100		70	104	39	64
13	66	93		114	117		24	61		59	91	99	111
14	45	78		6	26		93	111		53	92	35	67
15	42	61		12	45		43	84		54	91	30	49
16	86	104		54	96		15	52		109	116	96	107
17	58	78		50	84		21	63		30	52	55	75

	Research Activity			Student Support & Outcomes			Diversity			Overall Measure			
Program	Research Activity: 5th Percentile	Research Activity: 95th Percentile		Student Support & Outcomes: 5th Percentile	Student Support & Outcomes: 95th Percentile		Diversity: 5th Percentile	Diversity: 95th Percentile		R Rankings: 5th Percentile	R Rankings: 95th Percentile	S Rankings: 5th Percentile	S Rankings: 95th Percentile
18	16	39		51	91		53	86		53	86	27	54
19	41	66		57	94		14	58		33	53	45	72
20	50	71		67	100		63	96		36	65	57	77
21	42	73		9	44		15	56		56	84	33	65
22	61	82		112	116		34	77		77	106	88	102
23	6	19		17	54		109	115		41	94	12	26
24	28	37		79	102		4	33		19	50	35	53
25	64	96		62	89		18	67		33	63	77	98
26	95	113		38	95		116	117		111	117	102	112
27	76	97		100	112		72	101		104	116	112	116
28	101	112		3	27		19	66		58	97	78	95
29	58	82		10	38		108	114		57	80	49	73
30	67	85		27	60		80	102		73	106	71	89
31	76	95		26	59		88	106		69	106	84	100
32	44	67		41	82		22	71		21	32	38	63
33	108	115		100	113		39	73		97	112	111	115
34	14	29		38	81		37	80		10	24	19	28

	Research Activity			Student Support & Outcomes			Diversity			Overall Measure			
Program	Research Activity: 5th Percentile	Research Activity: 95th Percentile		Student Support & Outcomes: 5th Percentile	Student Support & Outcomes: 95th Percentile		Diversity: 5th Percentile	Diversity: 95th Percentile		R Rankings: 5th Percentile	R Rankings: 95th Percentile	S Rankings: 5th Percentile	S Rankings: 95th Percentile
35	9	20		49	85		48	88		8	14	11	21
36	31	44		51	91		8	50		26	45	29	48
37	2	4		3	18		50	91		4	8	2	4
38	4	10		1	6		94	108		11	21	4	9
39	2	3		2	14		64	94		2	5	2	3
40	100	113		101	113		4	40		81	104	101	110
41	35	62		33	77		69	100		36	57	30	56
42	85	107		22	54		1	13		74	104	87	103
43	13	29		6	31		7	37		25	43	12	22
44	20	30		52	94		99	111		11	27	21	28
45	42	60		25	56		13	56		30	49	35	54
46	7	13		48	83		44	88		8	17	8	16
47	67	93		69	101		27	72		64	94	71	92
48	23	43		4	22		23	64		51	82	25	44
49	43	74		43	89		2	14		30	72	30	54
50	92	105		99	113		2	24		51	81	99	107
51	63	79		40	82		34	78		27	48	56	74

	Research Activity			**Student Support & Outcomes**			**Diversity**			**Overall Measure**			
Program	Research Activity: 5th Percentile	Research Activity: 95th Percentile		Student Support & Outcomes: 5th Percentile	Student Support & Outcomes: 95th Percentile		Diversity: 5th Percentile	Diversity: 95th Percentile		R Rankings: 5th Percentile	R Rankings: 95th Percentile	S Rankings: 5th Percentile	S Rankings: 95th Percentile
52	87	101		30	86		3	23		75	107	68	87
53	109	116		2	22		95	110		69	100	77	95
54	20	36		13	43		37	85		16	28	18	27
55	116	117		47	85		14	54		91	112	112	116
56	72	101		48	97		6	40		81	113	73	98
57	85	103		34	80		22	70		54	89	90	102
58	51	69		70	101		58	90		39	78	65	81
59	100	115		114	117		46	85		45	81	102	112
60	114	117		100	112		110	115		73	105	114	117
61	84	98		30	68		5	33		51	79	78	94
62	24	40		55	94		98	113		34	63	36	63
63	35	61		58	90		18	61		52	82	37	68
64	39	66		10	33		97	109		55	85	36	66
65	98	110		6	27		1	4		76	107	74	92
66	47	87		12	37		23	61		76	112	33	68
67	69	90		23	60		72	97		38	71	50	71
68	62	100		8	34		100	112		87	107	55	82

	Research Activity			Student Support & Outcomes			Diversity			Overall Measure			
Program	Research Activity: 5th Percentile	Research Activity: 95th Percentile		Student Support & Outcomes: 5th Percentile	Student Support & Outcomes: 95th Percentile		Diversity: 5th Percentile	Diversity: 95th Percentile		R Rankings: 5th Percentile	R Rankings: 95th Percentile	S Rankings: 5th Percentile	S Rankings: 95th Percentile
69	77	90		26	69		18	64		29	58	57	75
70	49	70		90	107		18	59		31	53	53	74
71	103	115		102	113		64	97		109	117	115	117
72	85	100		42	93		36	73		61	108	73	91
73	68	91		18	47		10	48		55	86	52	75
74	40	63		88	107		75	100		54	81	49	76
75	32	45		28	72		12	52		37	63	27	42
76	23	36		35	76		19	64		6	19	21	29
77	100	115		103	114		8	46		107	116	107	113
78	91	105		52	97		28	75		74	105	89	102
79	32	53		33	73		13	53		60	87	37	63
80	100	114		29	78		6	43		85	108	86	102
81	51	67		17	49		29	74		38	67	40	57
82	9	16		9	36		69	101		5	11	7	12
83	43	60		70	97		37	83		11	25	40	60
84	36	49		51	98		12	58		41	65	29	44
85	24	48		80	103		59	98		26	49	30	63

	Research Activity			**Student Support & Outcomes**			**Diversity**			**Overall Measure**			
Program	Research Activity: 5th Percentile	Research Activity: 95th Percentile		Student Support & Outcomes: 5th Percentile	Student Support & Outcomes: 95th Percentile		Diversity: 5th Percentile	Diversity: 95th Percentile		R Rankings: 5th Percentile	R Rankings: 95th Percentile	S Rankings: 5th Percentile	S Rankings: 95th Percentile
86	61	82		101	111		9	45		29	55	76	92
87	13	26		71	102		6	35		8	19	12	21
88	82	100		21	55		36	88		81	106	78	94
89	67	90		41	74		66	98		51	75	70	90
90	21	31		97	109		2	21		16	59	27	38
91	50	68		60	95		68	98		32	58	50	69
92	87	104		114	117		36	74		70	105	104	112
93	104	115		27	68		13	42		106	117	102	114
94	4	7		48	85		94	107		3	7	5	9
95	57	76		36	78		116	117		23	36	57	77
96	62	89		22	53		38	78		75	101	55	81
97	91	114		33	88		6	48		64	98	90	105
98	59	90		61	99		2	23		87	112	60	86
99	11	26		1	7		102	111		13	27	8	17
100	13	22		8	31		47	88		5	10	10	18
101	30	46		84	105		58	88		32	52	31	52
102	8	16		26	63		6	45		7	17	9	17

	Research Activity			Student Support & Outcomes			Diversity			Overall Measure			
Program	Research Activity: 5th Percentile	Research Activity: 95th Percentile		Student Support & Outcomes: 5th Percentile	Student Support & Outcomes: 95th Percentile		Diversity: 5th Percentile	Diversity: 95th Percentile		R Rankings: 5th Percentile	R Rankings: 95th Percentile	S Rankings: 5th Percentile	S Rankings: 95th Percentile
103	14	22		73	100		85	104		15	28	19	25
104	6	10		3	19		110	115		3	12	5	8
105	1	1		2	12		82	103		1	1	1	1
106	5	11		4	28		87	106		28	60	5	9
107	68	93		63	100		11	52		62	99	90	104
108	31	44		51	89		42	80		30	56	29	44
109	71	103		1	11		35	76		105	116	59	88
110	36	48		29	69		4	36		32	51	30	45
111	17	74		102	113		25	71		80	112	31	91
112	43	72		15	46		85	108		69	98	44	75
113	15	25		8	28		59	98		9	17	10	17
114	35	60		36	74		20	68		20	35	27	49
115	15	24		9	36		65	100		18	29	14	21
116	52	70		55	94		112	115		41	74	48	67
117	37	53		7	42		61	96		25	49	26	37