This product is part of the RAND Corporation conference proceedings series. RAND conference proceedings present a collection of papers delivered at a conference. The papers herein have been commented on by the conference attendees and both the introduction and collection itself have been reviewed and approved by RAND Science and Technology.

CONFERENCE PROCEEDINGS

The U.S. Scientific and Technical Workforce

Improving Data for Decisionmaking

Terrence K. Kelly, William P. Butz, Stephen Carroll,
David M. Adamson, Gabrielle Bloom, editors

CF-194-OSTP/SF

June 2004

Prepared for the Office of Science and Technology Policy and the Alfred P. Sloan Foundation

RAND SCIENCE AND TECHNOLOGY

The research described in this report was conducted by the Science and Technology Policy Institute (operated by RAND from 1992 through November 2003) for the Office of Science and Technology Policy under Contract ENG-9812731 and the Alfred P. Sloan Foundation.

ISBN: 0-8330-3651-3

Published 2004 by the RAND Corporation
1700 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
201 North Craig Street, Suite 202, Pittsburgh, PA 15213-1516
RAND URL: http://www.rand.org/
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

Preface

About This Document

The U.S. scientific, technical, engineering, and mathematics (STEM) workforce makes key contributions to the nation's economic growth, national security, and other national goals. Given the importance of this workforce, monitoring and understanding its health and vitality are in the national interest. In 2003, a RAND Corporation study examined the issue of potential labor shortages in this workforce, which has been a recurring concern in federal policy circles since the 1950s. The study posed two questions: Are the current data on this workforce adequate to support relevant decisionmaking and, if not, what improvements are necessary?

To address this issue, the Office of Science and Technology Policy (OSTP) and the Alfred P. Sloan Foundation asked RAND to convene a technical conference to discuss the current state of data gathering on the U.S. STEM workforce and how data for decisionmaking might be improved. The conference included participants from federal research and development (R&D) and statistical agencies and researchers from universities and foundations. This volume provides each paper delivered at the conference, as well as three sections that RAND analysts prepared: an introduction, a rapporteur's summary, and list of priority data needs. The RAND materials have been peer-reviewed and edited. The conference papers, however, have not been peer reviewed and have been edited only for formatting and stylistic consistency.

The conference proceedings should be of interest to the science and technology policy community, science and math educators, students assessing career paths, and analysts interested in data and statistical issues.

Related RAND documents include

- William P. Butz, Gabrielle A. Bloom, Mihal E. Gross, Terrence K. Kelly, Aaron Kofner, and Helga E. Rippen, "Is There a Shortage of Scientists and Engineers? How Would We Know?" Santa Monica, Calif.: RAND Corporation, IP-241-OSTP, 2003. Online at http://www.rand.org/publications/IP/IP241/.
- William P. Butz, Terrence K. Kelly, David M. Adamson, Gabrielle A. Bloom, Donna Fossum, and Mihal E. Gross, *Will the Scientific and Technical Workforce Meet the Requirements of the Federal Government?* Santa Monica, Calif.: RAND Corporation, MG-118-OSTP, 2004. Online at http://www.rand.org/publications/MG/MG118/.

About the Office of Science and Technology Policy

The Office of Science and Technology Policy (OSTP) was created in 1976 to provide the president with timely policy advice and to coordinate the federal investment in science and technology.

About the S&T Policy Institute

Originally created by Congress in 1991 as the Critical Technologies Institute and renamed in 1998, the Science and Technology Policy Institute is a federally funded research and development center sponsored by the National Science Foundation. The S&TPI was managed by RAND from 1992 through November 30, 2003.

The Institute's mission is to help improve public policy by conducting objective, independent research and analysis on policy issues that involve science and technology. To this end, the Institute

- supports the Office of Science and Technology Policy and other Executive Branch agencies, offices, and councils
- helps science and technology decisionmakers understand the likely consequences of their decisions and choose among alternative policies
- helps improve understanding in both the public and private sectors of the ways in which science and technology can better serve national objectives.

In carrying out its mission, the Institute consults broadly with representatives from private industry, institutions of higher education, and other nonprofit institutions.

Inquiries regarding the work described in this document may be directed to the address below.

Debra Knopman
Assistant Director
RAND Science and Technology
1200 South Hayes Street
Arlington, VA 22202-5050
Tel: 703.413.1100, ext. 5667
Web: www.rand.org/scitech

Contents

Figures

Tables

Acknowledgments

The authors wish to thank the conference participants for their contributions to the event and to this document. We would also like to thank our reviewers, Charles Goldman and Roger Benjamin of RAND, for their thoughtful comments and questions, as well as Debra Knopman and Steven Popper of RAND S&T for their insights. We also owe thanks to Lisa Sheldone of RAND S&T for coordinating peer reviews and publication; to Phyllis Gilmore for her careful and rigorous editing; and, last but not least, to Mary DeBold for her tireless assistance in organizing the conference and in helping us produce this document.

Abbreviations

AA	associate in arts
ACS	American Communities Survey
AGEP	Alliance for Graduate Education and the Professoriate
B&B	Baccalaureate and Beyond
BEST	Building Engineering and Science Talent
BLS	Bureau of Labor Statistics
BPS	Beginning Postsecondary Students
BS	bachelor of science
CCLI	Course, Curriculum, and Laboratory Improvement
CED	Committee for Economic Development
CHERI	Cornell Higher Education Research Institute
CIP	Classification of Instructional Programs
CMS	Competency Management System
CNSTAT	Committee on National Statistics
CPS	Current Population Survey
CPST	Commission on Professionals in Science and Technology
DAS	Data Analysis Systems
DGE	Division of Graduate Education (NSF)
DHS	Department of Homeland Security
DNA	deoxyribonucleic acid
EHR	Education and Human Resources Directorate
ELS	Education Longitudinal Study of 2002
EPSCoR	The Experimental Program to Stimulate Competitive Research
EU	European Union
GAO	General Accounting Office
GK–12	Graduate Teaching Fellows in K-12 Education program (NSF)
GPRA	Government Performance Results Act (1993)
GRF	Graduate Research Fellowships (NSF)

GSS	Survey of Graduate Students and Postdoctorates in Science and Engineering
GUIRR	Government-University-Industry Research Roundtable
IERI	Interagency Education Research Initiative, a joint initiative by the National Science Foundation and the Department of Education to Support scientific research that studies educational interventions.
IGERT	Integrative Graduate Education and Research Traineeship program (NSF)
IPEDS	Integrated Postsecondary Education Data System
IT	information technology
ITAA	Information Technology Association of America
K–12	kindergarten through 12th grade
LED	Longitudinal Establishment Data
LSAMP	Louis-Stokes Alliances for Minority Participation
MBA	master of business administration
MORG	Merged Outgoing Rotation Group
MS	master of science
NAS	National Academy of Sciences
NASA	National Aeronautics and Space Administration
NBER	National Bureau of Economic Research
NCES	National Center for Education Statistics
NDEA	National Defense Education Act of 1958
NELS	National Education Longitudinal Study
NIH	National Institutes of Health
NPSAS	National Postsecondary Student Aid Study
NRC	National Research Council
NS&E	natural science and engineering
NSB	National Science Board
NSCG	National Survey of College Graduates
NSF	National Science Foundation
NSOPF	National Study of Postsecondary Faculty
NSRCG	National Survey of Recent College Graduates
OECD	Organisation for Economic Co-operation and Development
OMB	Office of Management and Budget
OSTP	Office of Science and Technology Policy
PhD	doctor of philosophy
R&D	research and development
R&RA	research and related activities
RETA	Research, Evaluation and Technical Assistance

ROLE	Research on Learning and Education
S&E	science and engineering
S&T	science and technology
S&TPI	Science and Technology Policy Institute
SBE	Social, Behavioral and Economic Sciences
SDR	Survey of Doctorate Recipients
SED	Survey of Earned Doctorates
SEI	Science and Engineering Indicators
SESTAT	Scientists and Engineers Statistical Data System (a comprehensive system of information about scientists and engineers in the United States, maintained by NSF)
SEVIS	Student and Exchange Visitor Information System
SHCP	Strategic Human Capital Plan
SOC	Standard Occupational Classification
SRS	Science Resources Statistics
STEM	scientific, technical, engineering, and mathematics
STEP	Science and Technology Expansion Program
SUNY	State University of New York
Thomson ISI	A data resource firm that provides information for scholarly research; a subsidiary of the Thomson Corporation
TPC	Teacher Professional Continuum
UIUC	University of Illinois at Urbana-Champaign
UNESCO	United Nations Educational, Scientific, and Cultural Organization
U.S.C.	United States Code
USSR	Union of Soviet Socialist Republics
WMPD	Women, Minorities and Persons with Disabilities in Science and Engineering

Part I
Prologue

CHAPTER ONE

Introduction

Overview[1]

Among many knowledgeable observers, the size and adequacy of the U.S. scientific, technical, engineering, and mathematics (STEM)[2] workforce have been recurring concerns. There are fears that the U.S. STEM workforce is aging and that its labor pool may soon dwindle. There are parallel fears that looming shortfalls in key skill areas may erode U.S. leadership in some science and engineering fields and that the growing proportion of non-U.S. citizens obtaining STEM degrees in the United States could complicate the task of mobilizing U.S. scientific and technological manpower for homeland security. However, evidence for these periodically anticipated shortages in the general STEM workforce has been hard to find. Indications of resulting national crises have, so far, been even less evident.

The failure of previously anticipated STEM workforce shortages to materialize should not be grounds for complacency. Were such shortages to arise, the implications could be serious, perhaps more so now than in earlier decades, because of the increasingly global demand for scientific and technical skills, the rise of formidable new competitors (such as China and India) in scientific and technological fields, and threats to homeland security. The continuing low entry rate of female and minority students into many STEM fields, noted particularly in National Science and Technology Council (2000) and National Science Board (NSB, 2003), also has serious implications. No less damaging for those affected are unemployment, underemployment, and clogged career tracks that greet new STEM graduates in fields when job growth is slow.

The implications of possible shortages and surpluses justify timely monitoring and examination of the STEM workforce in the United States. But the data needed to evaluate and assess STEM workforce trends and patterns and the available STEM jobs have often been inadequate for informed and timely policymaking.

In response to these data limitations, the Office of Science and Technology Policy (OSTP), in conjunction with the Sloan Foundation, asked the RAND Corporation to organize a blue-ribbon conference to identify the limitations of the available STEM workforce data and the opportunities for improving these data. The workshop brought leading

[1] This overview draws from recent RAND Corporation work, including Butz et al. (2003) and Butz et al. (2004).

[2] This workforce includes workers at all educational levels who perform functions in these fields (e.g., computer technicians whose highest formal degree is a high-school diploma, practicing engineers, medical doctors, and research scientists). It does not include those with degrees in STEM fields who are not currently working in STEM occupations.

researchers, science agency policymakers, and statistical agency experts together to address the following question: "How can we improve the data system for decisionmaking with respect to the U.S. STEM workforce?" This volume contains the proceedings of that conference, consisting of the papers delivered and discussed at the workshop, as well as RAND's synthesis of STEM workforce data needs and opportunities for meeting those needs.

STEM Workforce Shortages: A Recurring Concern

Alarms over the numbers of STEM personnel graduating and working in the United States have been raised on numerous occasions. The earliest alarm, triggered by Sputnik and accompanied by concerns about K–12 education in the United States, led to landmark federal legislation, the National Defense Education Act of 1958 (NDEA). The federal government dramatically increased funding for science and engineering education, resulting in the production of more newly trained scientists and engineers than there were jobs available during the early 1970s.

More recently, in the mid-1980s, the National Science Foundation (NSF) predicted "looming shortfalls" of scientists and engineers. No such shortfalls occurred and subsequent government hearings criticized the NSF for releasing these predictions.

Roughly a decade later, the Information Technology Association of America (ITAA) projected massive shortfalls in the availability of information technology workers. The Office of Technology Assessment (1988), relying on the ITAA analyses, echoed their warnings of future shortfalls of STEM workers. Again, there is no evidence that the predicted shortfalls occurred. A General Accounting Office (GAO) assessment of these projections criticized the methods used to develop them and the data on which they were based.

Various organizations have continued to examine the STEM workforce and have argued that their results imply growing gaps between the numbers of positions that require STEM workers and the numbers of STEM workers available. These organizations include the Institute of Medicine (1995), the National Research Council (2000a), and the National Science and Technology Council (2000).

In view of an "unfolding crisis for U.S. science and technology," a task force of the NSB called for "a coordinated response to meet our long-term needs for science and engineering skills in the U.S. workforce" (NSB, 2003, p. 1). The report went on to recommend a national policy imperative, stating that "all stakeholders must mobilize and initiate efforts that increase the number of U.S. citizens pursuing science and engineering studies and careers" (NSB, 2003, p. 2).[3]

What are the likely causes of these supposed shortfalls? Sputnik raised a red flag about the quality of American education and the need for a stronger federal role in improving STEM education in particular. More-recent concerns about the STEM workforce have arisen for other reasons. One has been the rapid emergence of new high-technology fields, such as information technology and genomics, fields that are deemed important for national competitiveness and security and that require technically trained workers. Another reason is the increasing technical demands of jobs that, even in traditional manufacturing and service

---

[3] Michael Crosby, NSB Executive Officer and NSB Office Director, also discussed the NSB report during the workshop (see Dr. Crosby's paper in Part II).

industries, call for workers with higher levels of technical training. Other reasons for concern have focused on the education pipeline—the gradually decreasing numbers of master of science degrees and doctorates awarded in several STEM fields, compared to the increasing requirements that are sometimes predicted.

Another related challenge, emphasized in the recent NSB report, is the long-term decline in the proportion of U.S. citizens among STEM doctorates that American institutions grant. It is argued that this trend has potentially ominous implications for the future of American leadership in science and engineering and that the trend could make it difficult to meet any increased demand for STEM professionals from national security and homeland security fields, whose duties require security clearances.

Diverse reasons have been offered for the decline in the proportion of U.S. citizens earning doctorates at U.S. institutions:

- Preparation for and interest in science careers in K–12 classrooms in the United States has declined.
- Recent long economic expansions have made the immediate postbaccalaureate job market increasingly attractive to potential graduate students.
- Careers in medicine, dentistry, veterinary medicine, law, business, and other professional fields, which compete with science for bright U.S. undergraduates, have become relatively more popular.
- Because the time invested in graduate and postdoctoral training has increased, so have the opportunity costs of preparation for entry into science careers.
- Demographic trends also have an effect—specifically, minority populations (such as Hispanics and African-Americans) that, for a variety of reasons, have traditionally been less likely to pursue STEM careers have grown.

Other possible causes for the decline in the number and proportion of U.S. citizens earning STEM doctorates have overseas roots. For example,

- The increasing attractiveness of holding a doctorate earned in the United States could lead more foreign students to apply to U.S. universities, thus displacing U.S. students.
- The international professional networks that have been established over decades have had a snowball effect, with a growing stream of top foreign candidates applying to work with their U.S. mentors and friends.
- Foreign science students increasingly desire to live and work in the United States after training here.[4]

However, many of these claims of shortfalls are suspect or are based on metrics that must be taken in context. Viewed broadly since the 1950s, evidence for the periodically anticipated shortages in the general STEM workforce has been hard to find. Indications of

---

[4] However, the number of foreign students applying to U.S. graduate schools has declined in the wake of visa rule changes following September 11, 2001. If this development becomes a long-term trend, it could affect the labor pool for the U.S. STEM workforce. See, for example, a recent survey conducted by the Council of Graduate Schools, which reported that graduate applications from international students have declined at more than 90 percent of U.S. universities for the fall 2004 term, and the number of submissions fell 32 percent from 2003 ("New Survey Confirms Sharp Drop in Applications to U.S. Colleges from Foreign Graduate Students," 2004). There have also been concerns that foreign students who would once have applied to U.S. universities might instead be choosing universities in the United Kingdom and Australia (Associated Press, 2003).

resulting national crises have, so far, been even less evident. Ironically, the closest thing to a crisis was perhaps the distress of unemployed and underemployed engineers in the early 1970s, mathematicians and physicists in the 1990s, molecular and cellular biologists in the late 1990s, and Silicon Valley scientists and engineers thereafter. But these are the manifestations of surpluses, not shortages, in the STEM workforce.

Statements about shortages based on such metrics as declining percentages of U.S. citizens earning doctorates must be viewed in context. These metrics do not use standard economic measures to assess the actual need but do shed some light on other issues of importance. For example, for national security reasons, it may be desirable to have more STEM workers who are U.S. citizens; for social reasons, it might be desirable for more minorities and women to earn STEM degrees. In both cases, the meaning of the word "shortage" must be clearly understood for these claims to make sense.

Finally, although previously anticipated STEM workforce shortages have not materialized in the economic sense, the implications of a shortage of skills critical to U.S. growth, competitiveness, and security justify continued examination of the nature and sources of the production of scientists, technical workers, engineers, and mathematicians in the United States, as well the demand for their services.

Improving the Data System for Decisionmaking

Data limitations have severely hampered past analyses of the STEM workforce and the design of appropriate policies for that workforce. The continually shifting definition of the STEM workforce, including whether it is best characterized by degree field or current occupation or job, compounds these difficulties. However the workforce is defined, the data available for evaluating the numbers of STEM graduates and workers and the number of jobs available for them have been inadequate for informed and timely policymaking. Finally, there has been little behavioral modeling and estimation of how these particular labor markets adjust to changes in supply and demand.[5] Some of these deficiencies are more damaging than others, depending on the particular focus of monitoring or analysis. Some are easier to correct than others, depending on whether the cause of difficulty is in the source of the data or in the aggregation and reporting.

The decentralized nature of the U.S. policymaking process—with many stakeholders at different levels of government, some public, some private—places a special premium on the sharing and transparency of data. Given such a process, better data may be the most effective way to improve the efficiency of the various relevant markets.

This conference attempted to address several of these data issues. In particular, it was intended to improve the range, quality, and timeliness of the data on the STEM workforce by inviting experts to identify specific data series whose (improved) collection and dissemination would substantially increase the ability of government policymakers, the private sector, university administrators, and students to recognize impending shortages or surpluses of STEM workers in particular fields. Leading labor market researchers, policymakers and administrators from federal science agencies, and administrators and technical experts from federal statistical agencies and the private sector participated in the workshop.

---

[5] The point is made in National Research Council (2000b).

The workshop consisted of five sessions: The first addressed whether the United States is indeed facing a shortage of scientists and engineers. The papers and subsequent discussion focused on problems associated with monitoring and forecasting STEM workforce conditions.[6]

The second featured leading labor market analysts who explored the kinds of data labor market researchers need. Using careful conceptual definitions of shortage, these analysts identified empirical measures that could stand as adequate proxies for the conceptual definitions and that federal statistical agencies or other data providers could produce in a timely manner.[7]

In the third, staff of federal technical agencies and other organizations that use STEM workforce data discussed what kinds of data science and technical policymakers need. They detailed the nature of the decisions they make that would benefit from improved data. At the end of this session, the workshop rapporteur listed the types of decisions that require improved data, as well as many crosscutting considerations that had arisen in the day's discussions of particular data improvements.

The fourth concentrated on the data needs that had emerged from earlier sessions. Federal statistical agency staff provided focus for this discussion via a presentation on meeting the data needs, outlining the opportunities and challenges for the U.S. government. Staff members identified ongoing and planned data-collection and analysis efforts that can meet the data needs identified in the earlier sessions. Furthermore, they cited particular data needs that current plans cannot meet because of budgetary, technical, or organizational reasons.

In the final session, the workshop rapporteur summarized the types of decisions driving requirements for more and better data, listed crosscutting considerations, detailed 40 specific data requirements and the prospects for their fulfillment, and suggested priorities for future data-gathering efforts.

Approximately 25 expert staff members from government agencies and private organizations participated actively in the discussions. Their informed questions, comments, and suggestions identified important considerations that would otherwise have passed unnoted and added to the practicality of the workshop's recommendations. For an extended discussion of the themes they raised, please see Part III, the Rapporteur's Summary.

Organization of This Document

This volume presents the conference papers in the order of their delivery. Many presenters prepared written papers, which are included in this report. Other presenters delivered their remarks accompanied by overhead presentations.

Following the individual presentations, a rapporteur's summary attempts to organize the workshop's major themes in a way designed to assist follow-up action. This section summarizes the decisions driving requirements for more and better data, lists crosscutting considerations, details specific data requirements, and identifies the prospects for satisfying each of these. The final chapter presents RAND's assessment of priorities for the federal statistical system and other data providers, as informed by workshop presentations and discus-

---

[6] For a list of speakers in all sessions, see Appendix B.

[7] Butz et al. (2003) explores several definitions of "shortage" at more length.

sion and by our own experience and judgment. These priorities reflect a balancing of data users' stated requirements against the budgetary, technical, and organizational factors that constrain data providers.

References

Associated Press, "International Student Enrollment Slows in U.S.," CNN.com, November 3, 2003. Online at http://www.cnn.com/2003/EDUCATION/11/03/foreign.students.ap/ (as of April 13, 2004).

Butz, William P., Gabrielle A. Bloom, Mihal E. Gross, Terrence K. Kelly, Aaron Kofner, and Helga E. Rippen, "Is There a Shortage of Scientists and Engineers? How Would We Know?" Santa Monica, Calif.: RAND Corporation, IP-241-OSTP, 2003 (available online at http://www.rand.org/publications/IP/IP241).

Butz, William P., Terrence K. Kelly, David M. Adamson, Gabrielle A. Bloom, Donna Fossum, and Mihal E. Gross, *Will the Scientific and Technical Workforce Meet the Requirements of the Federal Government?* Santa Monica, Calif.: RAND Corporation, MG-118-OSTP, 2004 (available online at http://www.rand.org/publications/MG/MG118).

Institute of Medicine, *Reshaping the Graduate Education of Scientists and Engineers*, Washington, D.C.: National Academy Press, 1995.

National Research Council, *Forecasting the Demand and Supply of Scientists and Engineers*, Washington, D.C.: National Academy Press, 2000a.

_____, Office of Scientific and Engineering Personnel, *Measuring the Science and Engineering Enterprise: Priorities for the Division of Science Resources Studies*, Washington, D.C.: National Academy Press, 2000b.

National Science and Technology Council, *Ensuring a Strong U.S. Scientific, Technical, and Engineering Workforce in the 21st Century*, Washington, D.C.: Office of Science and Technology Policy, April 2000. Online at http://www.ostp.gov/html/00411_3.html (as of June 2, 2004).

National Science Board, Committee on Education and Human Resources, Task Force on National Workforce Policies for Science and Engineering, *The Science and Engineering Workforce: Realizing America's Potential*, Washington D.C., NSB 03-69, 2003.

"New Survey Confirms Sharp Drop in Applications to U.S. Colleges from Foreign Graduate Students," *Chronicle of Higher Education*, March 4, 2004. Online at http://chronicle.com/prm/daily/2004/03/2004030403n.htm (as of June 2, 2004).

Office of Science and Technology Policy, Congress of the United States, *Educating Scientists and Engineers: Grade School to Grad School*, Washington, D.C., June 1988.

Part II
Contributed Papers

The following papers were delivered at the conference on the U.S. STEM workforce in December 2003. Except where noted, these papers have not undergone peer review.

CHAPTER TWO

Do We Need More Scientists?

Michael S. Teitelbaum[1]

For much of the past two decades, predictions of an impending shortage of scientists and engineers in America have gained increasingly wide currency. The country is failing to produce scientists and engineers in numbers sufficient to fulfill its economic potential, the argument runs. The supposed causes are weaknesses in elementary, secondary, or higher education, inadequate financing of the fields, declining interest in science and engineering among American students, or some combination of these. Thus it is said that the United States must import students, scientists, and engineers from abroad to fill universities and work in the private sector—though even this talent pool may dry up eventually as more foreign nationals find attractive opportunities elsewhere.

Yet alongside such arguments—sometimes in the very same publications in which they appear—one learns of layoffs of tens of thousands of scientists and engineers in the computer, telecommunications, and aerospace industries, of the deep frustration and even anger felt by newly minted PhDs unable to find stable employment in traditional science and engineering career paths, and of senior scientists and engineers who are advising undergraduates against pursuing careers in their own fields. Why the contradictory reports on professions routinely deemed critical to the success of the American economy? Is it possible that there really is no shortage in these fields?

A History of Gloomy Forecasts

Pronouncements of shortages in American science and engineering have a long history. They date at least to the late 1950s, around the time the USSR launched Sputnik, the first orbiting satellite, prompting concerns that an era of Soviet technological advantage over the United States had emerged. The United States responded with massive public investments in science and engineering education. This led to sharp increases in the numbers pursuing such studies and a surfeit in the 1970s of entry-level scientists and engineers.

The recent history of shortage forecasts begins in the mid-1980s, when the then-leadership of the National Science Foundation (NSF) and a few top research universities began to predict "looming shortfalls" of scientists and engineers in the next two decades

---

[1] Michael S. Teitelbaum is a demographer and program director at the Alfred P. Sloan Foundation, in New York. This paper was originally published in *The Public Interest*, National Affairs, Inc., No. 153, Fall 2003, Washington, D.C. Reprinted with permission. The views expressed here are those of the author and do not necessarily represent those of the Alfred P. Sloan Foundation.

(Atkinson, 1990).[2] Their arguments were based upon quite simplistic demographic projections produced by a small policy office reporting to the NSF director—projections that earlier had been sharply criticized by the NSF's own science and engineering workforce experts.[3]

Only a few years later, it became apparent that the trends actually pointed toward a growing surplus of scientists and engineers. In 1992, the House Committee on Science, Space and Technology's Subcommittee on Investigations and Oversight conducted a formal investigation and hearing about the shortfall projections, leading to much embarrassment at the NSF. In his opening remarks at the hearing, the subcommittee's chairman, Democrat Howard Wolpe of Michigan, declared that the "credibility of the [National Science] Foundation is seriously damaged when it is so careless about its own product." Sherwood Boehlert, the subcommittee's ranking Republican and now chair of the full House Science Committee, called the NSF director's shortfall predictions "the equivalent to shouting 'Fire' in a crowded theater." They were "based on very tenuous data and analysis. In short, a mistake was made," he said. "Let's figure out how to avoid similar mistakes, and then move on." (U.S. House of Representatives, 1993, pp. 1–10.)

Boehlert's advice was not heeded. Only five years later, during the high-tech boom of the late 1990s, an industry association known as the Information Technology Association of America (ITAA) began to produce a series of reports asserting burgeoning gaps and shortages of information-technology workers, based on proprietary surveys of what it termed "job openings." The first ITAA report claimed that some 190,000 information-technology jobs could not be filled in 1997 (ITAA, 1997).[4] The second concluded that there were 346,000 open positions in 1998. The Department of Commerce then produced its own report, which drew heavily upon the findings of the two ITAA reports.

The General Accounting Office (GAO) published a sharply critical assessment of these three related reports in 1998. It concluded that all their shortfall estimates were questionable due to the studies' weak methodologies and very low response rates. Unabashed, ITAA returned to the fray in 2000. Its third report asserted that over 843,000 information-technology positions would go unfilled that year due to a shortfall of qualified workers. Despite withering criticism from the GAO, the ITAA reports provided useful political support for the successful lobbying campaign for dramatic expansion—to the current level of 195,000 per year—of the H-1B visa, the temporary-visa program for the foreign "specialty workers" that constitute the bulk of foreign science and engineering professionals being admitted to work in the United States.

Remarkably, even the recent economic downturn does not seem to have deterred proponents of the workforce shortage theory. Take NASA administrator Sean O'Keefe, who invoked a shortage argument during testimony before the House Science Committee in October 2002 on NASA's hiring problems. "Throughout the Federal government, as well as

---

[2] The article was based upon Dr. Atkinson's President's Lecture to the Annual Meeting of the American Association for the Advancement of Science, 18 February 1990, New Orleans. Additional accessible reports on these materials may be found in, e.g., Holden (1989) and Bloch (1990), p. 25.

[3] This was a small staff office located within the NSF Director's office. The 1992 congressional investigation described below uncovered extensive documentary evidence, reproduced in the subcommittee report, that NSF's own professional experts on the science and engineering workforce had expressed strong skepticism about the validity of the shortfall projections.

[4] This study was conducted, mysteriously enough, by a staff member of the Cato Institute, known for its libertarian ideology rather than its labor market research.

the private sector, the challenge faced by a lack of scientists and engineers is real and is growing by the day," O'Keefe told the committee.

The following month a new organization called Building Engineering and Science Talent (BEST) published a report entitled The Quiet Crisis: Falling Short in Producing American Scientific and Technical Talent. This "quiet crisis," notes Jackson (2002),

> stems from the gap between the nation's growing need for scientists, engineers, and other technically skilled workers and its production of them. ... This "gap" represents a shortfall in our national scientific and technical capabilities.

Some business leaders and academics are also advancing the shortage thesis despite the economic downturn. Two reports with findings similar to the BEST study subsequently emerged in the spring of 2003. One was a report addressed to the Government-University-Industry Research Roundtable (GUIRR) of the National Academies (Jackson, 2003), and the other was prepared by the Committee for Economic Development (CED), an organization of business and education leaders.

Even some associated with the NSF seem unchastened by the embarrassing failure of the "shortfall" projections of a decade ago. In June 2003, the National Science Board, the NSF's governing body, released for public comment a draft task-force report addressing the "unfolding crisis" in science and engineering. "Current trends of supply and demand for [science and engineering] skills in the workplace indicate problems that may seriously threaten our long-term prosperity, national security, and quality of life," it said.

The Evidence

The profound irony of many such claims is the disjuncture between practice in the scientific and engineering professions—in which accurate empirical evidence and careful analyses are essential—and that among promoters of "shortage" claims in the public sphere, where the analytical rigor is often, to be kind, quite weak. Few, if any, of the market indicators signaling shortages exist. Strong upward pressure on real wages and low unemployment rates relative to other education-intensive professions are two such indicators conspicuously absent from the contemporary marketplace.

A RAND study released in 2003 assembled the available data from its own research, the NSF, the Census Bureau, the Bureau of Labor Statistics (BLS), the National Research Council (NRC), and several scientific associations. What RAND found largely discredits the case being made for labor shortages. First, RAND noted the obsolescence of the available data, the newest of which refers mostly to 1999 or 2000. RAND called this "especially unfortunate" given that "the [science and engineering] workforce situation has arguably changed significantly" since the heady times of the dot-com, information technology, and telecom booms. But more importantly, RAND's analysis of data even from the boom period showed that "neither earnings patterns nor unemployment patterns indicate [a science and engineering] shortage in the data we were able to find" (Butz et al., 2003, p. 4).

Recent government unemployment data tend to confirm these findings. Data for the first and second quarters of 2003 released by the Bureau of Labor Statistics showed surprisingly high unemployment rates in science and engineering fields. Even the recently "hot" computer and mathematical occupations are experiencing unemployment of 5.4 to 6 percent. For computer programmers, the numbers range from 6.7 to 7.5 percent. All engineering (and architecture) occupations taken together are averaging 4.4 percent unemployment, while the rates for the high-tech fields of electrical and electronic engineering are in the range of 6.4 to 7 percent. Reported unemployment in the life, physical, and social sciences ranges from 2.8 to 4.1 percent. Many of these numbers are remarkably high for such high-skill occupations. Unemployment for the whole of the U.S. workforce averaged about 6 percent over the same period, and highly educated groups, such as scientists and engineers, normally have substantially lower unemployment rates than the national average (BLS, 2003).

In the natural-science disciplines, which employ far fewer people than engineering, numerous reports by leading scientists have been pointing to increasingly unattractive career prospects for newly minted PhDs. As one example among many, a 1998 National Academy of Sciences (NAS) committee on careers in the life sciences—the largest field in the natural sciences—reported that "recent trends in employment opportunities suggest that the attractiveness to young people of careers in life-science research is declining" (NRC, 1998, p. 1). More recent data from 2002 showed that key indicators of career problems had continued to deteriorate since then, prompting Shirley Tilghman, the NAS committee's chair and current president of Princeton University, to tell Science magazine that she found the 2002 data "appalling." She said the data reviewed earlier by the committee looked "bad" at the time, "but compared to today, they actually look pretty good" (Goldman and Marshall, 2002, p. 40).

The 2003 RAND study concurred. Butz et al. (2003, p. 4) concluded that

> Altogether, the data ... do not portray the kind of vigorous employment and earnings prospects that would be expected to draw increasing numbers of bright and informed young people into [science and engineering] fields.

It is of course quite possible to have "appalling" early career problems in some areas of science and engineering alongside very good career prospects in others. Administrators of federal technical agencies, such as NASA, do face special problems, such as hiring freezes or other ongoing personnel or financial constraints. Senior personnel at NASA and other agencies have been offered substantial early retirement incentives, while hiring procedures to replace them tend to be cumbersome and slow. In "hot" fields that are new or growing rapidly, like bioinformatics, human resources are inevitably in short supply. And truly exceptional scientists and engineers will always be few in number and vigorously pursued by employers.

Still, in most areas of science and engineering at present, the available data show sufficient numbers or even surpluses of highly qualified candidates with extensive postgraduate education. This is especially the case in the academy, which has become risk-averse about replacing departing tenured faculty with tenure-track junior positions. Instead, many universities in the United States have been filling such open slots with temporary and part-time appointees they find in ample pools of highly educated applicants. Indeed, advertisements for a single tenure-track assistant professorship often attract hundreds of applications from recent PhDs. Similar circumstances prevail for engineers and scientists in large sectors of the U.S. economy, such as telecommunications, computing, and software, sectors in which lurching market collapses and large bankruptcies have greatly weakened demand for their services.

What Does the Future Hold?

Many recent shortage claims point not to current circumstances, but to projections of future demand. What can be said with reasonable assurance about such predictions?

Unfortunately, labor-market projections for scientists and engineers that go more than a few years into the future are notoriously difficult to make. An expert workshop convened by the National Research Council in 2000 reported universal dissatisfaction with past projection efforts, and stated declaratively that "accurate forecasts have not been produced" (NRC, 2000).

The workshop report commented in particular upon one such study that is often cited by shortage proponents: the Bureau of Labor Statistics' "Occupational Outlook." The most recent "outlook," completed in 2001, projected that over the next decade computer-related fields, including software engineers, computer-network and system administrators, and analysts, would likely be the fastest growing occupations nationwide. But the NRC workshop report noted the limitations inherent in such projections (NRC, 2000, pp. 28–29):

> The omission of behavioral responses makes the BLS outlook unreliable as a basis for decisions on federal funding designed to respond to anticipated shortages. ... The BLS outlook neglects many dimensions in which adjustment may occur, including training and retraining, and especially in response to changes in wages. ... No response is built into time trends in relative occupational wages on either the demand side (where employers substitute capital for labor when relative wages rise) or the supply side (where students move toward occupations in which relative wages are rising).

One might add that many science and engineering fields are heavily influenced by federal funding, which makes projections of future workforce demand dependent upon quite unpredictable political decisions and world events. To their credit, the authors of the BLS Occupational Outlook themselves emphasize the need for caution. "The BLS projections were completed prior to the tragic events of September 11 ... [and] the nature and severity of longer-term impacts [of the terror attacks] remains unclear," the authors write. "At this time, it is impossible to know how individual industries or occupations may be affected over the next decade."

Owing to such events and unforeseeable changes in the market, no one can know what the U.S. economy and its science and technology sectors will look like in 2010. It follows that no credible projections of future "shortages" exist on which to base sensible policy responses.

Misdirected Solutions

Not only are claims of current or future shortages inconsistent with all available quantitative evidence, but alas many of the solutions proposed to deal with the putative "crisis" are profoundly misdirected. The most popular proposed solutions seem to focus mainly on the supply side, urging action to increase the numbers of U.S. students pursuing degrees in science and engineering. Recommendations often include calls for reform of the U.S. elementary and secondary education systems, especially inadequacies in science and mathematics; informational efforts to promote knowledge of such careers among U.S. secondary school students and of the science and math prerequisites required to pursue them at university level; financial and other incentives to increase interest in such fields among U.S. students; and increases in the number of "role models" in science and engineering fields for women and underrepresented minorities. Other commentators, apparently more pessimistic about the abilities of U.S. students, recommend increasing the numbers of students or workers from abroad to meet the needs of the U.S. economy.

This focus on supply to the virtual exclusion of demand is not warranted. However desirable many of these proposals may be on other grounds, they are unlikely to be very effective in attracting U.S. students to careers in science and engineering unless employment in these fields is sufficiently attractive to justify the large personal investments needed to enter them. Surprisingly enough, it is far from common to hear this rather obvious point raised in public discussions of the subject. To put the matter more succinctly, those who are concerned about whether the production of U.S. scientists and engineers is sufficient for national needs must pay serious attention to whether careers in science and engineering are attractive relative to other career opportunities available to American students. And yet little such attention has been forthcoming in recent years.

The qualifications for careers in engineering and especially in science involve considerable personal investments. The direct financial costs of higher education in the sciences can be staggering, depending on the financial circumstances of undergraduates and their families, whether the institution is private or public, whether postbaccalaureate education is required, and whether such education is subsidized.

Engineering and science differ substantially in these characteristics. For engineering, only the baccalaureate is normally required for entry into the profession. Most engineering B.S. degrees are earned at state universities, which are heavily subsidized by state governments. In addition, direct financial aid is often available for those in financial need. In contrast, professional careers in the sciences now commonly require completion of the PhD and increasingly require subsequent postdoctoral work. The direct financial costs of this extensive graduate and postdoctoral work are typically heavily subsidized by both government and universities. Yet even with such subsidies, the personal costs to qualify as a scientist can be quite high—mainly due to the lengthening time required to attain the degree and complete postdoctoral training.

The extreme case is that of the biosciences, which account for half of all PhDs awarded in the natural sciences. Over the past couple of decades, the average period of required postbaccalaureate study has increased dramatically, to between nine and twelve years from about seven to eight years. The PhD itself has stretched out to seven or eight years from about six, while the now-essential postdoctoral apprenticeship has lengthened to between two and five years from one or two in decades past. In career terms, this means that most young bioscientists cannot begin their careers as full-fledged professionals until they are in their early thirties or older, and those in academic positions usually are not able to secure the stable employment that comes with tenure until their late thirties. Unsurprisingly, the idea of spending nine to twelve years in postbaccalaureate studies before one is qualified for a real job may be unattractive to substantial numbers of would-be young scientists.

There are also concerns about negative impacts on scientific creativity. Wendy Baldwin, until recently the deputy director for extramural research at the National Institutes of Health (NIH), notes concerns arising at NIH over "the long-held observation that a lot of people who do stunning work do it early in their careers" (Goldman and Marshall, 2002, p. 40). Bruce Alberts, in his 2003 President's Address to the National Academy of Sciences, described as "incredible" the fact that even though NIH funding has doubled in only the past five years, the average age of first-time grant recipients has continued increasing. "Many of my colleagues and I were awarded our first independent funding when we were under 30 years old ... [now] almost no one finds it possible to start an independent scientific career under the age of 35," Alberts told the academy (Alberts, 2003). Nobel laureate and codiscoverer of DNA structure James Watson agrees. As he put it in characteristically pithy terms in a 1992 interview,

> I think you're unlikely to make an impact unless you get into a really important lab at a young age. ... People used to be kings when they were nineteen, generals. Now you're supposed to wait until you're relatively senile.

It's not hard to see why this also portends ill for science careers at a personal level. Delaying career initiation until one's thirties poses inherent conflicts with marriage and family life. Many who might be attracted to careers in science are justifiably concerned that such a career choice comes at too high a personal cost.

The problem has not gone unnoticed. Many scientific societies have decried the trend toward longer degrees and postdoctoral apprenticeships, and U.S. universities have created more than 70 new two-year graduate science degrees designed for those who wish to pursue scientific careers outside of the academy. (Start-up costs of many of these have been supported by Sloan Foundation grants.) These new degrees, called "Professional Science Master's degrees," have been attracting interest among U.S. science majors who might otherwise choose paths leading to business or law school (see Sloan Foundation, undated).

Opportunity Costs

Some senior scientists stress that no one should pursue a science career to get rich, which is a point well taken. Yet it would be unwise simply to ignore how alternative career paths compare in strictly economic terms. The nine- to twelve-year period that a would-be bioscientist now must spend in a student role or a low-paid postdoctoral position means that a substantial fraction of lifetime income that would otherwise be earned must be foregone.[5] This is what economists term opportunity costs, and these are by no means insignificant. A 2001 study conducted by a team of leading economists and biologists for the American Society for Cell Biology found that bioscientists experience a "huge lifetime economic disadvantage" on the order of $400,000 in earnings discounted at 3 percent compared to such PhD fields as engineering, and about $1 million in lifetime earnings compared with medicine. When expected lifetime earnings of bioscientists are compared with those of MBA recipients from the same university, the study's conservative estimates indicate a lifetime earnings differential of $1 million, exclusive of stock options. When stock options are included, the differential doubles to $2 million (Freeman et al., 2001, pp. 10–12).

[5] Plus deferred benefits such as pension contributions, which with tax-free accumulation can become very significant sums over time.

In smaller scientific fields, such as physics and chemistry, where PhD programs are shorter and lengthy postdoctoral work less universal, the differentials are smaller but still substantial. Given the direct financial costs and opportunity costs, careers in science and engineering must offer significant attractions relative to other career paths available to American students. College graduates with demonstrated talent and interest in science and mathematics can choose to go to medical school, law school, or business school; they can pursue other professional education; or they can enter the workforce without graduate degrees.

The options available to most foreign students—at least for those from poorer countries—are completely different. Most do not have the option to study at U.S. medical, law, or business schools (due to the high costs and lack of subsidies), nor can they easily enter the U.S. workforce directly. In contrast, science PhD programs at many American universities actively recruit and subsidize graduate students and postdoctoral fellows from China, India, and elsewhere, from which positions many are able to move on to employment in the United States.

There are, of course, many significant noneconomic rewards associated with careers in science and engineering: the intellectual challenge of research and discovery, the life of the mind in which fundamental puzzles of nature and the cosmos can be addressed, and the potential to develop exciting and useful new technologies. For some, these attractions make science and engineering careers worthy of real sacrifices—they are "callings" rather than careers, analogous to those of religious or artistic vocations. Happily, a number of talented students will decide, based on personal values and commitments, to pursue graduate degrees and careers in science or engineering, even with full knowledge that the career paths may be unattractive in relative terms. Yet it is also true that others with strong scientific and mathematical talents will decide that a better course for their lives would be to go directly into the workforce rather than to follow additional scientific studies, or instead to pursue professional degrees in business, law, or other fields.

The Politics of Shortages

Public discourse about these issues is mired in paradox. There are energetic claims of "shortages" of engineers, while unemployment rates are high and mid-career engineers face increasing job instability. There are reprises of earlier "shortage" claims about scientists, while undergraduates demonstrating high potential in science and math increasingly seem to be attracted to other careers. Some emphasize the need for K–12 reform, even though very large numbers of entering college freshmen intend to major in science or engineering but later choose otherwise. The NIH research budget has doubled within only a few years, but the average age at which scientists win their first research grants is rising. Why are shortage claims so persistent despite so much evidence to the contrary?

On this issue, where one stands depends upon where one sits. Most of the assertions of current or impending shortages, gaps, or shortfalls have originated from four sources: university administrators and associations, government agencies that finance basic and applied research, corporate employers of scientists and engineers and their associations, and immigration lawyers and their associations.

The economist Eric Weinstein has uncovered documentary evidence suggesting that the real intent of some of those involved in the 1980s "shortfall" alarms from NSF may have been to limit wage increases for PhD scientists (Weinstein, undated). Whether or not such motivations underlay that episode, we can certainly appreciate the various incentives that may currently spur some to endorse such claims. Universities want to fill their classrooms with undergraduates who pay their fees and finance their research with external funding and, to do so, recruit graduate students and postdoctoral fellows to teach undergraduates and to staff their research laboratories. Government science-funding agencies may find rising wages problematic insofar as they result in increased costs for research. Meanwhile, companies want to hire employees with appropriate skills and backgrounds at remuneration rates that allow them to compete with other firms that recruit lower-wage employees from less affluent countries. If company recruiters find large numbers of foreign students in U.S. graduate science and engineering programs, they feel they should be able to hire such noncitizens without large costs or lengthy delays. Finally, immigration lawyers want to increase demand for their billable services, especially demand from the more lucrative clients, such as would-be employers of skilled foreign workers.

None of these groups is seeking to do harm to anyone. Each finds itself operating in response to incentives that are not entirely of its own making. But a broad commonality of interests exists among these disparate groups in propagating the idea of a "shortage" of native-born scientists and engineers. Moreover, claims of shortages in these fields are attractive because they have proven to be effective tools to gain support from American politicians and corporate leaders, few of whom would claim to be experts on labor markets. As noted earlier, the dubious reports from the ITAA were used successfully to convince the Congress to triple the size of the H-1B visa program in 2000. In late 2002, a leading lobbyist for the National Association of Manufacturers, responding to criticism that shortage claims cannot be supported by credible evidence, put the matter succinctly: "We can't drop our best selling point to corporations," he explained.

Such a short-term view is naturally attractive to lobbyists because it works politically. But it may turn out to be a serious error over a longer period. Claims of impending shortages can easily become self-fulfilling prophecies if, as in the past, government responds by subsidizing education or increasing visas for foreign workers without seriously considering the effects such actions may have upon the attractiveness and sustainability of career paths for such professionals. Action along these lines could create an even larger surfeit of scientists and engineers—one that drives down the number of Americans willing to enter these professions and, paradoxically, creates the very problem it seeks to address.

Instead of raising the false flag of shortages, those concerned about the future of science and engineering in the United States should encourage objective appraisals of current career paths, as well as innovations in higher and continuing education designed for more agile adjustments to inevitable changes in these dynamic fields. The overarching goal should be to find ways to make these careers attractive relative to the alternatives, for this is the only sustainable way to ensure a supply commensurate with the United States' science and engineering needs.

References

Alberts, Bruce, "Harnessing Science for a More Rational World," 2003 President's Address, 140th Annual Meeting, National Academy of Sciences, Washington, D.C., April 28, 2003.

Atkinson, Richard, "Supply and Demand for Scientists and Engineers: A National Crisis in the Making," *Science*, Vol. 248, April 27, 1990, pp. 425–432.

Bloch, Erich, Director, National Science Foundation, testimony before U.S. Senate, Committee on Commerce, Science and Transportation, Subcommittee on Science, Technology, and Space, Hearing on Shortage of Engineers and Scientists, May 8, 1990.

Bureau of Labor Statistics, Current Population Survey, Table 3, Employed and experienced unemployed persons by detailed occupation and class of worker, Quarter I, 2003, unpublished.

Butz, William P., Gabrielle A. Bloom, Mihal E. Gross, Terrence K. Kelly, Aaron Kofner, and Helga E. Rippen, "Is There a Shortage of Scientists and Engineers? How Would We Know?" Santa Monica, Calif.: RAND Corporation, IP-241-OSTP, 2003 (online at http://www.rand.org/publications/IP/IP241).

Freeman, Richard, Eric Weinstein, Elizabeth Marincola, Janet Rosenbaum, and Frank Solomon, "Careers and Rewards in Bio Sciences: The Disconnect Between Scientific Progress and Career Progression," manuscript submitted to the American Society for Cell Biology, September 2001.

Goldman, Erica, and Eliot Marshall, "NIH Grantees: Where Have All the Young Ones Gone?" *Science*, Vol. 298, October 4, 2002, p. 40.

Holden, Constance, "Wanted: 675,000 Future Scientists and Engineers," *Science*, June 30, 1989, pp. 1536–1537.

Information Technology Association of America, "Help Wanted: The IT Workforce Gap at the Dawn of a New Century," Washington, D.C., 1997.

Jackson, Shirley Ann, *The Quiet Crisis: Falling Short in Producing American Scientific and Technical Talent*, Washington, D.C.: Building Engineering and Science Talent (BEST), September 2002.

______, *Envisioning a 21st Century Science and Engineering Workforce for the United States: Tasks for University, Industry, and Government*, Washington, D.C.: National Academies Press, 2003.

National Research Council, *Forecasting Demand and Supply of Doctoral Scientists and Engineers: Report of a Workshop on Methodology*, Washington, D.C.: National Academy Press, 2000.

Sloan Foundation (sponsor), Professional Science Masters, Web page, undated. Online at http://www.sciencemasters.com (as of June 3, 2004).

U.S. House of Representatives, Committee on Science, Space, and Technology, Subcommittee on Investigations and Oversight, *Projecting Science and Engineering Personnel Requirements for the 1990s: How Good Are the Numbers?* Washington, D.C.: U.S. Government Printing Office, 1993, pp. 1–10.

Watson, James, interview at the "Winding Your Way Through DNA" symposium at the University of California, San Francisco, Washington, D.C.: The National Health Museum, 1992. Online at http://www.accessexcellence.org/AB/CC/watson.html (as of June 3, 2004).

Weinstein, Eric, "How and Why Government, Universities, and Industry Create Domestic Labor Shortages of Scientists and High-Tech Workers," National Bureau of Economic Research, undated. Online at http://nber.nber.org/~peat/PapersFolder/Papers/SG/NSF.html (as of June 3, 2004).

CHAPTER THREE

What Will It Take for the United States to Maintain Global Leadership in Discovery and Innovation?

Michael P. Crosby and Jean M. Pomeroy[1]

Abstract

The National Science Board (NSB) is a strong advocate of the need for better data to inform policy on the science and engineering (S&E) workforce, and to guide both individual career decisions and program development. A number of NSB's policy reports have recommended improving data for policy and planning for the workforce. The most recent is *The Science and Engineering Workforce: Realizing America's Potential* (NSB, 2003). This report reiterates earlier concerns about the adequacy of data for S&E workforce policy, which had been raised in NSB reports on graduate and postdoctoral education, and on setting priorities for research in the federal budget. The board does not believe there is an immediate crisis of supply and demand in S&E. Rather, the United States is in a long-distance race to retain its essential global advantage in S&E human resources and sustain our world leadership in science and technology. The board's policy recommendations, therefore, focus on what is needed for the United States to retain an S&E workforce that enables it to sustain long-term global preeminence in discovery and innovation. A high-quality, diverse, and adequately sized workforce that draws on the talents of all U.S. demographic groups and talented international students and professionals is crucial to our continued leadership, and is therefore a vital federal responsibility.

Introduction

NSB—in particular, the Chair of the Board, Dr. Warren Washington; the Vice Chair of the Board, Dr. Diana Natalicio; the Chair of the NSB Committee on Education and Human Resources, Dr. George Langford; and the Chair of the NSB Task Force on National Workforce Policies for Science and Engineering, Dr. Joseph Miller—were pleased to have the opportunity to participate in this important forum, sponsored jointly by the White House Office of Science and Technology Policy and Sloan Foundation.

---

[1] Michael P. Crosby is an executive officer and Jean M. Pomeroy is a senior policy analyst with the National Science Board, Arlington, Virginia. The authors would like to acknowledge the valuable contributions to this paper by the Division of Science Resources Statistics, National Science Foundation.

By way of background, NSB was established by Congress in 1950 (see NSB, 2004). It has 24 members appointed by the president and confirmed by the Senate, plus the National Science Foundation (NSF) director as an ex-officio member. The board has dual responsibilities:

- to oversee and guide the activities of, and establish policies for, the NSF
- to serve as an independent national S&E policy body that provides advice to the president and the Congress on policy issues related to science and engineering, identified by the president, Congress, or the board itself.

As part of its national S&E policy responsibility, the board prepares its biennial report on science and engineering indicators. However, it also periodically reports on an ad hoc basis on important issues affecting science and engineering research and education. This paper will refer primarily to the board's most recent report (NSB, 2003) in presenting the board's perspectives on the broad topic of this forum and Michael Teitelbaum's "Do We Need More Scientists?" (see Chapter Two).

Data Needs for Policy and Planning

The data needs Teitelbaum mentions have also been identified by the board in its recent policy reports as critically needed information for S&E workforce policy and program planning (NSB, 1996, 1997, 2001, and 2003). These include the need to better:

- generate high-quality, more timely data to guide policy
- study paths to S&E careers
- identify and publicize strategies that make S&E careers more attractive to students
- understand the impacts of the international S&E workforce on U.S. S&E capabilities
- understand domestic supply and demand for S&E skills, with immediate special focus on doctorates and postdoctorates
- study time to degree, duration of postdoctoral appointments, and transition rates to independent careers for graduate students and new PhDs.

An Appropriate Focus for a National Dialogue

The board feels that perhaps a more appropriate question is not "do we need more scientists?" but "what will it take for the United States to maintain global leadership in discovery and innovation?" The board's recent report underscores that "The United States is in a long-distance race to retain its essential global advantage in S&E human resources and sustain our world leadership in science and technology (NSB, 2003, p. 41)." A high-quality, diverse, and adequately sized workforce that draws on the talents of all U.S. demographic groups and talented international students and professionals is crucial to our continued leadership and is therefore a vital federal responsibility. No one has the perfect recipe for achieving our goal for the national S&E workforce. This workshop provided a useful forum to help explore the possibilities for measuring and estimating our national needs for the future.

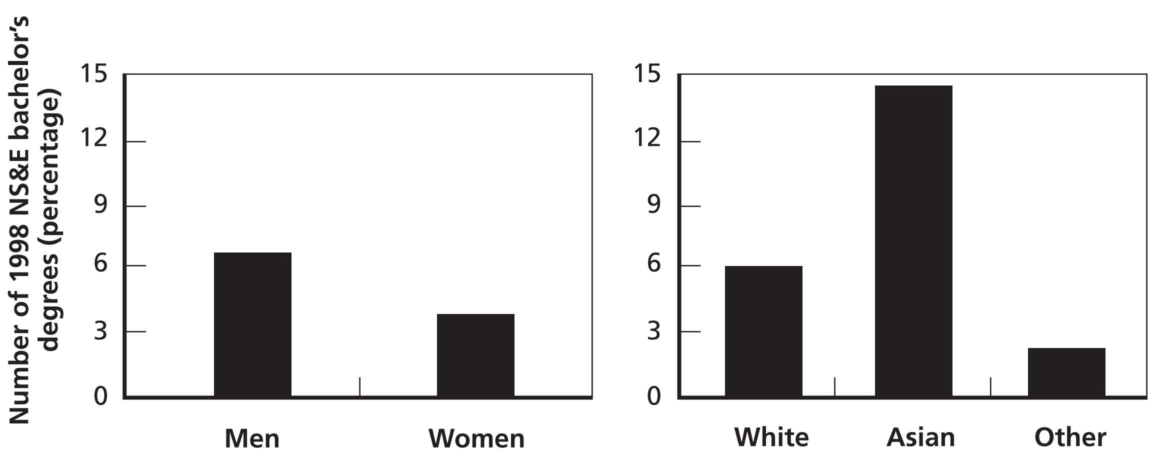
Briefly, the board's concerns about the U.S. ability to sustain its S&E workforce second to no other nation are based on a combination of a number of clear and unfavorable trends. S&E jobs in the United States have grown faster than the overall workforce for a long time; however,

- There is a lack of growth in the number of bachelor's degrees earned by U.S. citizens in natural science and engineering (NS&E) fields.
- Long-term demographic trends show increasing shares of the college-aged population will be from groups that are underrepresented in NS&E.
- U.S. dependence on scientists and engineers born in other countries is increasing at all degree levels.
- Global competition for S&E talent is growing.
- The U.S. S&E workforce is aging.

Figure 3.1 shows that women and underrepresented minorities earn NS&E bachelor's degrees in numbers far below what would be expected by their shares of the population. Men are much more likely than women to earn a baccalaureate in NS&E fields—at a rate of 7.5 percent versus 4.6 percent.[2] For ethnic groups, whites and Asians far exceed Hispanics, blacks and Native Americans in their participation rates in NS&E fields—6.3 percent and 14.7 percent, respectively, for whites and Asians versus 2.6 percent for underrepresented minorities.

Participation in the U.S. workforce by scientists and engineers with a bachelor's, master's, or PhD, and born in foreign countries has increased greatly over the last decade (Figure 3.2) from 14 percent to 22 percent of S&E positions. The proportion of the S&E workforce with PhDs who are foreign born is nearly 40 percent of the most highly educated workers.

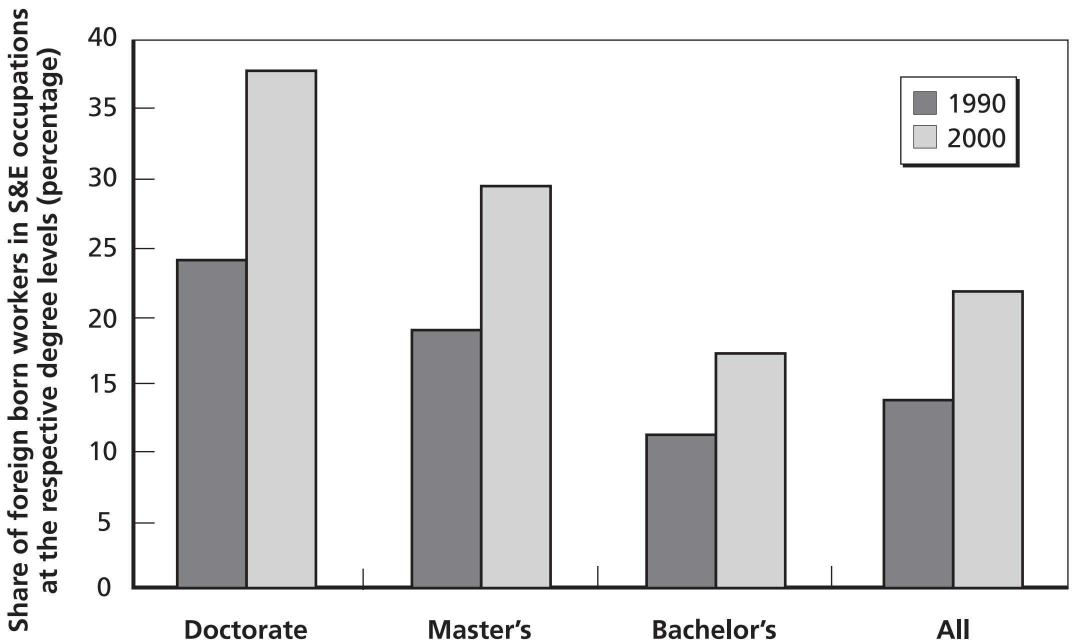
Figure 3.1
Women and Underrepresented Minorities Are Less Likely to Earn Bachelor's Degrees in NS&E



SOURCE: Science and Engineering Indicators 2002, Text Table 2-9, page 2-23, Appendix table 2-18.
NOTE: Percentages are number of 1998 NS&E bachelor's degrees to 100 of the 24-year-old population.
RAND *CF194-3.1*

[2] Percentages are the ratio of 1998 bachelor's degrees to 100 of the 24-year old population (NSB, 2002, Vol. 2, Appendix table 1-18).

Figure 3.2
Foreign-Born Workers Account for an Increasing Share of the U.S. S&E Workforce



SOURCE: U.S. Bureau of the Census, 5 percent public use microdata system files, 1990 and 2000.
RAND *CF194-3.2*

While these statistics and trends are alarming, the board recognizes a need for expanding the base of information about the entire S&E workforce system, from status and trends data acquisition to improving counseling and guidance on education pathways for S&E careers. Existing data sources have a number of limitations for informing federal policy and planning. Other presentations at this forum discussed the specifics concerning data related to the S&E workforce (Lynda Carlson, Director, Science Resources Statistics, NSF, concerning data related to the S&E workforce and Judith Ramaley, Assistant Director, Education and Human Resources, NSF, on S&E education). Therefore, this paper will not go into any detail on these topics.

The board has recommended that the federal government should lead a national effort to build a base of information in a number of specific areas, including

- status of the S&E workforce
- S&E skill needs and utilization
- strategies that attract high-ability students and professionals to S&E careers.

Common Definitions

One concern of the board is that dialogue on policy for the S&E workforce needs to be based on a common set of definitions. The board's definition of the S&E workforce is broader than those being employed by some others who are commenting on this important topic. From the perspective of sustaining a world leading S&E workforce, it is important to include consideration of all individuals educated in S&E fields and using their skills in their jobs, not just those officially classified as a "scientist" or "engineer."

Therefore the board's report emphasizes the need to take not only the long view with respect to our S&E workforce but also to look broadly at all levels of education. Indeed, the average scientist or engineer in the workforce has a baccalaureate in science or engineering.

Moreover, of those educated in S&E fields but not working in an officially designated science or engineering occupation, the vast majority report that they are using the S&E skills in their jobs. As Figure 3.3 shows, analyses that focus exclusively on commonly designated S&E occupations miss a substantial portion of the workforce with S&E degrees and using these S&E skills in their jobs. They are not factored into models and policymaking discussions.

For example, a scientist with an MS in physics may change careers from being a bench scientist in industry to a high school teacher of advanced placement physics. In doing so, she would go from a strictly defined S&E occupation to a non-S&E occupation. Yet, her new position is an absolutely vital component of the S&E enterprise in this country.

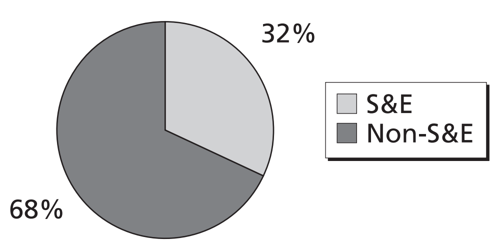
Another example that may apply to many of the participants at this workshop is an individual who has moved from a classic research occupation to science policy and administration. Both the physics teacher and science policy administrator are still using their S&E skills. Indeed, these skills are essential for high-level accomplishment in these positions. Therefore, NSB suggests that a broad definition of what constitutes the S&E workforce should be utilized for national policy purposes.

Global S&E

Much better data are needed to support U.S. policy on the international flow of S&E students and workers. This is an immediate and critical issue for U.S. S&E, given our growing dependence on international students and professionals.

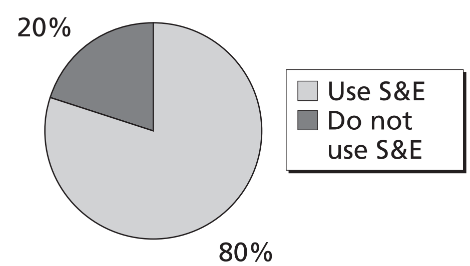
Figure 3.3
NSB Defines the S&E Workforce Broadly

Occupations of S&E workers



SOURCE: Science and Engineering Indicators 2002, Appendix Table 3-2, Page A3-5.
NOTE: SESTAT definitions of "S&E" and "Non-S&E" occupations.

Use of S&E skills in occupations



SOURCE: Calculated from Science and Engineering Indicators 2002 Text Tables 3-1, 3-2, pages 3-6, 3-7.
NOTE: "Use S&E skills" includes all those in SESTAT-defined "S&E" jobs and those in SESTAT-defined "Non-S&E" jobs who "closely" or "somewhat" use S&E skills in those jobs.

RAND *CF194-3.3*

NSB recommends that the federal government should substantially raise its investment in research that advances the state of knowledge on international S&E workforce dynamics. Monitoring the flow of international students and workers would require collaboration across federal and international agencies, including, for example, the Bureau of Labor Statistics, Census Bureau, NSF, Department of Homeland Security, and the Organisation for Economic Co-operation and Development.

Summary and Conclusion

NSB has identified some important trends that suggest the need for attention to future challenges to U.S. global preeminence in discovery and innovation, which is critically dependent on our large, diverse, and high-quality S&E workforce.

Available data on trends and limited studies can guide our S&E workforce policy discussions. However, better information and data are needed in areas of

- demand and supply for skills in the workforce
- understanding and publicizing career paths for scientists and engineers
- successful strategies to make S&E studies and careers more attractive and accessible
- the dynamics of the international S&E workforce.

NSB does not believe that there is an immediate crisis with the U.S. S&E workforce. However, the trends are clear and disturbing. The nation cannot afford to wait for a crisis or until our data sources are absolute before taking action to counter observed negative trends. Indeed it is unlikely, given the unpredictability of discovery and innovation, that we would ever be able to anticipate rapid changes as a result of paradigm shifting discoveries. Nor is it likely we can anticipate the workforce impacts of unique events in history, such as the end of the Cold War and the events of September 11.

While striving to improve data acquisition and analyses for planning and policy decisions regarding the future U.S. S&E workforce, inability to perfectly forecast near-term specific skill needs should not be an excuse for inaction in our strategic response to clear evidence that we do have on broad S&E trends. The federal government must continue to play a crucial role in developing our nation's human resources for S&E in order to sustain our global leadership over the long term.

References

National Science Board, *Report from the Task Force on Graduate and Postdoctoral Education*, Washington, D.C., NSB 96-02, 1996.

_____, *The Federal Role in Science and Engineering Graduate and Postdoctoral Education*, Washington, D.C., NSB 97-235, 1997.

_____, *Federal Research Resources: A Process for Setting Priorities*, Washington, D.C., NSB 01-156, 2001.

_____, *Science and Engineering Indicators 2002*, Washington, D.C., NSB 02-1, 2002.

_____, *The Science and Engineering Workforce: Realizing America's Potential*, Washington, D.C., NSB 03-69, Arlington, Va.: National Science Foundation, 2003.

______, Web site, last updated: May 19, 2004. Online at http://www.nsf.gov/nsb (as of June 3, 2004).

Teitelbaum, M., "Do We Need More Scientists?" *Public Interest*, Fall 2003.

CHAPTER FOUR

Does America Face a Shortage of Scientists and Engineers?

Ronald Ehrenberg[1]

Michael Teitelbaum's paper "Do We Need More Scientists?" (see Chapter Two in this volume) provides a provocative start to this meeting. The notion of shortages or surpluses existing in markets in which prices are free to adjust is somewhat alien to economists because, ultimately, price changes will bring markets into equilibrium. At best, concern might be expressed over the length of time it takes a market to adjust; in situations in which there are long lags in the response of supply to prices (such as in the production of PhDs), policies might be needed to facilitate the adjustment (such as *temporarily* changing the number of government-sponsored assistantships, fellowships, and traineeships provided for PhD students).

As someone who served on the committee that issued the 1998 study of the early careers of life scientists that Teitelbaum talks about in his article and who has critiqued models that projected shortages of new PhDs, I am very sympathetic to many of the points that he makes (National Research Council, 1998; Ehrenberg, 1991). What I want to focus on today is the word we in his title, because, as Teitelbaum emphasizes, the question of shortages or surpluses is often in the eye of the beholder. For example, from the perspective of faculty members involved in the academic enterprise, increased research project budgets lead to increased demand for graduate research assistants and postdoctoral fellows. Each faculty member wants to maximize his own research output, and concern about future employment prospects for one's students often falls by the wayside.

From the perspective of an academic institution, budget situations dictate the extent to which the institution has the resources to bid for top new faculty prospects or is forced to settle for lesser quality faculty members whose lower salaries it can afford. Most American college students are educated at public institutions, and hence, most American faculty members are employed at public higher education institutions. Over the last 25 years, the budget problems faced by state governments, coupled with the increased demand on their budgets for expenditures in areas other than higher education, have led state appropriations per student in public higher education to decline relative to tuition levels at private higher education institutions. Percentage increases in tuition levels at public higher education institutions have been roughly the same as those at private higher education institutions; however, because the publics started at a lower absolute level of tuition, their increases have not been

[1] Ronald G. Ehrenberg is the Irving M. Ives Professor of Industrial and Labor Relations and Economics at Cornell University in Ithaca, New York, and a research associate at the National Bureau of Economic Research, Washington, D.C. The author is grateful to the Andrew W. Mellon Foundation and the Atlantic Philanthropies (USA) Inc. for their financial support of CHERI. He emphasizes, however, that all thoughts expressed here are strictly his own.

large enough to permit their expenditure levels to rise at the same rate as expenditure levels in private higher education.

As a result, full-time faculty salaries have declined substantially in public higher education relative to the salaries of faculty in private higher education (Ehrenberg, 2003). In addition, a growing tendency to substitute part-time and non–tenure-track full-time faculty for full-time tenure and tenure-track faculty has occurred. Most people are unaware of the magnitude of these shifts, but at some campuses they have been enormous. For example the share of undergraduate credit hours generated by full-time tenured and tenure-track faculty members declined from 81 percent to 58 percent at the State University of New York (SUNY) University Centers (doctorate-granting institutions) between the fall of 1992 and the fall of 2001; the comparable decline at the SUNY University Colleges (master's-granting institutions) was from 84 percent to 70 percent (Ehrenberg and Klaff, 2003, table 2). The attractiveness of public higher education institutions, as potential employers for new PhD students, has declined, and the voluntary turnover of existing faculty at public institutions is now higher than that of their faculty counterparts at private institutions (Nagowski, 2003).

Along with the growing dispersion of resources between public and private higher education has come a growing dispersion of resources across private higher education institutions. Fueled by growing dispersions of endowment wealth caused by changes in stock market levels over the last 25 years and the tendency of the richest institutions to devote more of their annual giving to further building their endowments than do the poorer institutions, there has been a growing dispersion of average faculty salaries in private higher education (Ehrenberg, 2003; Ehrenberg and Smith, 2003).

Another byproduct of the growing dispersion of wealth and the efforts by universities to attract the best possible new faculty members to their ranks has been an escalating competition for top scientists and engineers that is manifested in large start-up cost packages. A survey conducted by the Cornell Higher Education Research Institute (CHERI) in the spring and summer of 2002 of science and engineering departments at our nation's research and doctoral universities found that start-up cost packages for new assistant professors at private research universities were typically in the $400,000 to $500,000 range, while packages at public universities were somewhat lower (Ehrenberg, Rizzo, and Condie, 2003).[2] So in spite of what one might consider a surplus of new PhDs in some science and engineering fields, the "price" needed to attract the best candidates is high. Because private universities more often have access to endowments and annual giving streams from which they can obtain funds for these start-up cost packages, it is not surprising that the public universities, more often than private universities, reported to us that they obtained at least part of the funding for their start-up cost packages by keeping positions vacant until salary savings can be achieved to cover the start-up costs. To the extent that institutions face a continual need to attract new faculty, this suggests a further permanent reduction in the size of the full-time tenure track faculty at many public institutions.

Is this trend, especially at public institutions, of substituting part-time and non–tenure-track full-time positions likely to continue in the future and thus, in the context of Teitelbaum's paper, to further reduce the attractiveness of PhD study in the sciences? To the extent that governors and state legislatures are concerned more about the undergraduate degrees that are generated by their public higher education institutions and less about these

[2] Start-up cost packages for senior faculty members are considerably larger and often exceed one million dollars.

institutions' faculty members' research, I fear that the answer will be yes, unless researchers can demonstrate that these shifts in faculty composition are having adverse effects on undergraduate students (such as increasing dropout rates and/or slowing time to degree). After all, from the perspective of an economist, substituting cheaper for more-expensive inputs and achieving the same output is very rational.

However, saying that state governments are behaving rationally is not saying that their actions are socially optimal from the perspective of the broader "we." Paul Romer and others have argued that the rate of growth of productivity in our economy depends upon the rate of growth of innovation in the economy, which in turn depends upon the share of our nation's educated workforce with degrees in science and engineering (Romer, 2000). So even if labor markets are in balance, there may not be a socially optimal number of American citizens and permanent residents pursuing science education.

Even if state government officials realize that income growth in their states depends upon the share of their residents that are educated in science and engineering, it does not follow that states are irrational in cutting back on their expenditures on public higher education. Recent research suggests that the proportion of the adult population in a state that is college educated is only very loosely tied to the expenditures that state governments are currently making on their public higher education systems (Bound et al., 2001). Mobility of college-educated workers across areas moderates the linkage between a state's expenditures on higher education and the composition of its adult workforce.

So to the extent that it is socially optimal to have more people trained as scientists and engineers in our workforce, how do we accomplish this? My reasoning above suggests that, ultimately, it is the federal government that must play the role of guaranteeing that "we" generate an adequate supply of scientists and engineers. However, there are roles for individual academic institutions to play.

For example, the changing structure of grading at many selective American colleges and universities, which has led average grades in the humanities and soft social sciences to rise relative to average grades in the sciences and economics, is surely a problem—it is easier for students to "do well" in nonscience disciplines, and this discourages them from pursuing science careers (Sabot and Wakeman-Linn, 1991; Parekh, 2002). Institutional policies that would require all classes (of a given level) to assign the same median grade, as is done in some law schools, might help. It might also help to provide more information on the nature of careers in science and engineering, on the fact that many individuals trained in science and engineering often wind up in management positions, and on the earnings of people who choose such careers.

Finally, some people have pointed to the growing share of PhD degrees in science and engineering being granted to foreign nationals as evidence that foreign students are crowding out potential U.S. students and have argued that limitations should be placed on their admission. As Teitelbaum points out, the growth in foreign enrollments is a logical response by American universities to the declining interest of American students in PhD study in the sciences and engineering. The only study that I know of that looked at the preferences of American universities for foreign graduate students found that the universities "discriminated" against foreign students. More precisely, American citizen student applicants had a higher probability of being admitted to doctoral programs than did foreign applicants with the same admission credentials (test scores) (Attiyeh and Attiyeh, 1997). So if we are concerned about not having enough PhDs in science and engineering, making it easier for

PhD graduates of our nation's universities who come from foreign countries to stay and work in the United States should be a desirable policy.

References

Attiyeh, Gregory, and Richard Attiyeh, "Testing for Bias in Graduate School Admissions," *Journal of Human Resources*, Vol. 32, Summer 1997, pp. 524–548.

Bound, John, Jeffrey Groen, Gabor Kezdi, and Sarah Turner, "Trade in University Training: Cross-State Variation in the Production and Use of College-Educated Labor," Cambridge, Mass.: National Bureau of Economic Research, working paper 8555, October 2001.

Ehrenberg, Ronald G., "Academic Labor Supply," in Charles T. Clotfelter, Ronald G. Ehrenberg, Malcolm Getz, and John J. Siegfried, *Economic Challenges in Higher Education*, Chicago, Ill.: University of Chicago Press, 1991.

_____, "Studying Ourselves: The Academic Labor Market," *Journal of Labor Economics*, Vol. 21, April 2003, pp. 267–287.

Ehrenberg, Ronald G., and Daniel B. Klaff, "Changes in Faculty Composition Within the State University of New York System: 1985–2001," Ithaca, N.Y.: Cornell Higher Education Research Institute, working paper 38, August 2003. Online at http://www.ilr.cornell.edu/cheri (as of June 3, 2004).

Ehrenberg, Ronald G., Michael J. Rizzo, and Scott S. Condie, "Start Up Costs in American Research Universities," Ithaca, N.Y.: Cornell Higher Education Research Institute, working paper 33, March 2003. Online at http://www.ilr.cornell.edu/cheri (as of June 3, 2004).

Ehrenberg, Ronald G., and Christopher L. Smith, "The Sources and Uses of Annual Giving at Selective Private Research Universities and Liberal Arts Colleges," *Economics of Education Review*, Vol. 22, June 2003, pp. 223–235.

Nagowski, Mathew, "Associate Professor Turnover at American Colleges and Universities," Ithaca, N.Y.: Cornell Higher Education Research Institute, working paper 31, August 2003. Online at http://www.ilr.cornell.edu/cheri (as of June 3, 2004).

National Research Council, *Trends in the Early Careers of Life Scientists*, Washington D.C.: National Academy Press, 1998.

Parekh, Jay, "Do Median Grades Vary Across Departments?" Ithaca, N.Y.: Cornell Higher Education Research Institute, working paper 30, December 2002. Online at http://www.ilr.cornell.edu/cheri (as of June 3, 2004).

Romer, Paul M., "Should the Government Subsidize Supply or Demand in the Market for Scientists and Engineers," Cambridge Mass.: National Bureau of Economic Research, working paper 7723, June 2000.

Sabot, Richard, and John Wakeman-Linn, "Grade Inflation and Course Choice," *Journal of Economic Perspectives*, Vol. 5, Winter 1991, pp. 59–170.

CHAPTER FIVE

Data! Data! My Kingdom for Data!
Data Needs for Analyzing the S&E Job Market

Richard B. Freeman[1]

> [T]o support development of effective science and engineering workforce policies and strategies, the Federal Government must: substantially raise its investment in research that advances the state of knowledge on international science and engineering workforce dynamics; lead a national effort to build a base of information on:
>
> 1. The current status of the S&E workforce
> 2. National S&E skill needs and utilization and
> 3. Strategies that attract high-ability students and professionals to S&E careers.
>
> —*National Science Board, 2003*

In 2003, the National Science Board (NSB), which advises the president and Congress on science policy issues, and serves as the governing board for the National Science Foundation (NSF), issued the newest of a seemingly endless stream of reports from organizations concerned with the health of the science and engineering (S&E) workforce. The NSB report recognized that young scientists and engineers faced serious difficulties in the job market in the 1990s but nonetheless came out with the standard recommendation that the United States should encourage more young persons into S&E careers.[2] It noted the increasing reliance of the country on the foreign-born in the nation's S&E workforce. And, as quoted above, the report made greater investment in research on the S&E job market an imperative, to allow the federal government to make more-rational, evidence-based policies than in the past.

This is not the first time a major organization concerned with U.S. science policy has declared that the country needs greater information and knowledge about the S&E workforce (Citro and Kalton, 1989). There has long been a relative paucity of timely information on this workforce to inform public discourse and policy deliberations. When I was a graduate student, John Dunlop pointed out that the United States has more and better data on agricultural products than it has on the most skilled workers in the country. While the base

[1] Richard B. Freeman is Ascherman Professor of Economics at Harvard University in Cambridge, Massachusetts, and Co-Director of the Labor and Worklife Program at the Harvard law School, and Director of the Labor Studies Program at the National Bureau of Economic Research, Washington, D.C. He is also Co-Director of the Centre for Economic Performance at the London School of Economics and Visiting Professor at the London School of Economics.

[2] Following NSB practice, I use the S&E terminology inclusively, so that mathematicians and technicians are covered. The term scientific, technical, engineering, and mathematics (STEM) has been adopted by many recently for this workforce.

of information on scientists and engineers has improved in recent decades, data on agriculture produce is still more plentiful than data on the S&E workforce.

NSB is right that more-effective policies toward S&E workers require a better base of information. But limited data are not the only problem in developing evidence-based policies toward the S&E workforce. Data can be useful for monitoring developments in the job market for scientists and engineers, but data are most useful when they illuminate the way the market operates in ways that let us model better how supply and demand will respond to potential policy interventions. A horse would not have won Bosworth Field for King Richard III. What he needed was a better battle plan based on a more-accurate model of how others were likely to respond to his actions upon gaining the crown.

This paper has two themes. The first is that while we need more and better data—improved counts of numbers of scientists and engineers along various dimensions—our greatest need is for data specifically designed to answer key analytic questions about the functioning of the labor market for the S&E workforce. The second theme is that labor market analysis offers an essential guide to the data and research necessary to accomplish the goals of NSB. To understand the nation's needs for S&E workers (and, by inference, the perpetual claims of shortages of scientists and engineers), to develop strategies that attract high-ability U.S. students to S&E, and to assess the costs and benefits of such strategies, we need rigorous labor market analyses.

Data Use Determines Data Needs

Why do we want better data on the S&E workforce? There are five basic reasons:

1. *To monitor developments in this workforce.* At the least, good monitoring alerts government agencies, employers, and others concerned with the well-being of science to potential problems in the market for the key input—highly trained S&E workers. Most monitoring involves surveys that document the number of scientists and engineers with diverse characteristics and the number of students enrolled or degrees granted in the fields. Two key issues highlighted by these data in recent years are the shift in supply from U.S.-born and permanent resident scientists and engineers to foreign-born scientists and engineers and the underrepresentation of women and some minorities in S&E careers. We must have reliable numbers to assess developments in these areas. In addition, we need good data on the "price" side of the market (broadly defined)—salaries, quality of jobs held, unemployment rates, and the like—to judge how S&E occupations are faring relative to other occupations in the economy that highly able students may find attractive.
2. *To understand the determinants of the supply of persons choosing scientific and engineering careers.* Behind supply numbers are individual decisions to choose careers in S&E as opposed to other areas. Since there is a long lag between training in S&E and entry into the job market, young persons must implicitly or explicitly forecast how they will fare years into the future. They must assess their own skills relative to the skills needed in S&E and also form some notion of the future job market in those careers. This invariably involves considerable uncertainty, in part because of the vagaries of government research initiatives and the way research fields can shift from hot to cold, or conversely. Foreign students must decide whether or not to come to the United States to study in S&E fields and then whether or not to remain in the United States to work. Foreign S&E graduates

must decide whether or not to immigrate to the United States to pursue their careers. And the existing stock of S&E workers must decide where to work and on what problems to work. In all cases, analysis of supply involves more than simple counts: It requires research into decisionmaking and the factors that influence decisions.

3. *To understand the changing demand for S&E workers.* The critical information on the demand side relates to the direction and magnitude of shifts in demand schedules. Almost all analysts agree that standard projection models do not provide accurate forecasts (National Research Council, 2000), which highlights the need to understand the factors that change demands for scientists and engineers. Among the factors that have altered demand in past years are the shift of scientific and engineering activity from universities to the business sector; decreased spending on the space program; the doubling of National Institutes of Health (NIH) spending; the national nanotech initiative; the bioshield initiative; the growth of large-scale team science, which creates new career patterns; and sending S&E work off-shore to lower-cost foreign sites, where the response of demand to cost considerations is critical. If President Bush makes placing a man on the Moon and Mars a national goal, spending at NASA will skyrocket, shifting demand for S&E workers in a very different direction than it would be absent such an initiative. In all these cases, analysis of demand involves more than simple counts; it requires research into the factors that influence decisions about research and development (R&D) spending and the use of S&E workers in different sectors.

4. *To understand the market dynamics that govern the response of the job market to changes in government policies.* From 1998 to 2003, NIH doubled its spending on health-related research. In most labor markets, such a huge change in demand would have greatly advanced the career prospects of young persons and drawn large numbers into the field. Yet in this period, NIH reported the virtual disappearance of scientists under the age of 35 from research grant recipients; one National Academy of Sciences committee reported a "crisis in career expectations," and another recommended no further increases in supply.[3] For their part, postdocs, largely in the biosciences, organized the National Postdoc Association to improve what they viewed as poor job conditions.

   Why did the doubling of research spending fail to improve (and possibly worsen) the career opportunities of young scientists? As best I can tell, the reason is that the market for research is structured so that increased spending goes to existing faculty researchers, who hire teams of graduate students and postdocs from around the world to do the bench work. Increased R&D augments the supply of young researchers in the U.S. market without creating (many) additional tenure-track academic positions for those young researchers to move into after their postdoctoral work. We need information on how the institutional details of supporting research affect the supply and demand for young researchers, on the impact of team production on scientific career paths, and on the employment practices of labs, to assess how government initiatives or other exogenous changes affect the S&E job market.

5. *To evaluate the contribution of S&E workers to economic progress and other national goals.* The government supports scientists and engineers doing basic research for two reasons.

[3] Committee on National Needs for Biomedical and Behavioral Scientists; Committee on Dimensions, Causes, and Implications of Recent Trends in the Careers of Life Scientists.

First, advances in knowledge are the fundamental cause of economic growth and rising living standards. Second, the private sector does not have the incentive to fund (much) basic research. The private sector will not invest the socially optimal amount of basic research because it cannot appropriate much of the gains. Basic research requires a long time perspective. The results are hard to appropriate, even with a strong patent system. And basic research is very risky: Some lines of research may have commercial payoffs, but others will not. Econometric analyses of company and industry R&D indicate that industrial R&D has a high payoff but with high variance. It is more difficult, however, to pin down the returns to basic research, much less the link from government support of basic research to ideas and findings to outputs. Would the basic research have been done without public support—say in another country? How dependent were the ideas that went into a product on the basic research? Ideally, a government that optimized national output would decide how much basic research to support on basis of expected social returns from the new products, processes, and services that industry will create from the research. Assessing the social returns of S&E work requires information on the production of knowledge from funded research and on the flow of ideas to business and eventual useful products.

The United States has several surveys that gather useful data for the analysis of the S&E job market. NSF's Scientists and Engineers Statistical Data System (SESTAT) provides information on the employment experience of S&E workers; the Survey of Doctorate Recipients (SDR) offers longitudinal information on careers; the Survey of Earned Doctorates (SED) has data on the population of PhDs granted in a given year; and the Survey of Industrial R&D gives information on spending and employment in industry. There are also general government surveys, such as the Census of Population and the Current Population Survey (CPS), which provide data on scientists and engineers, along with workers in other areas, and Bureau of Labor Statistics surveys of workers, such as the occupational employment survey, which provide information on scientists and engineers, as well as on workers in other occupations. Various agencies have administrative data on research grants, stipends, and other activities that can provide particular insight into how students and faculty use government funding. From outside the government, professional societies and other groups provide further information, including regular surveys of graduating classes in particular disciplines, of college placement offices on salaries, and so on (the Commission on Professionals in Science and Technology is one example).

Papers in this conference proceedings and elsewhere delineate the characteristics of these surveys better than I can. Along with other users of the surveys, I am highly appreciative of the information they provide, but I will concentrate in this essay on their weaknesses in meeting the five goals listed above and in possible directions for improvement.

Data Gathering for Monitoring

Current data gathering has been targeted largely at counting the number and characteristics of S&E workers to monitor developments in the S&E job market. The principal tools for doing this are large surveys based on complete populations (SED) or on samples derived from sizable national bases (SDR), or past Censuses. By seeking to produce definitive "official statistics," these regular surveys do not in general inform newly emergent policy concerns in a timely fashion. This has the unfortunate consequence that it opens the door for less-accurate and potentially flawed and biased data to play a role in policy debate, as occurred during the debate over expanding H1-B visas to meet alleged shortages of information technology workers. Had government statistics provided sound measures of how firms used H1-B visas, the number of budgeted vacant positions in the alleged shortage areas, and the extent to which U.S. residents had the "shortage" skills, the federal government would have been in a more competent position to make an informed decision about the visas.

Absence of timely accurate data on the economic situation of postdoctorate workers has been detrimental to policymaking in other areas as well. Some analysts believe that the economic problems of postdoctorate workers discourage young Americans from science careers, but we lack good information on the numbers of postdocs; their pay, benefits, and career aspirations; and what happens when they leave their postdoctorate job. The 1997 SDR had a special section on postdocs, but its next postdoctorate worker module will be in 2007, which makes it hard to determine whether the economic status of these scientific workers in the interim and thus to measure and devise policies to remedy their low pay, lack of medical benefits, inadequate protection of employee rights, and so on. In 2004, Sigma Xi undertook a national survey of postdoctorate workers, which will provide more timely data and ask more pointed questions than NSF asked in its postdoc module. But this is far from a substitute for a more frequent NSF survey of this critical population.

There have also been problems in getting accurate and timely counts of particular types of scientists and engineers. NSF had only one survey counting the number of foreign-trained scientists and engineers working in the United States in the 1990s: the National Survey of College Graduates, which was based on the 1990 Census. The SED counts foreign-born graduates from U.S. universities annually but provides no information on graduates from foreign institutions who immigrate to the United States. Table 5.1 shows that, by using the 1990 data on foreign-trained graduates as its population base in SESTAT, NSF significantly underestimated the foreign-born share of scientists and engineers at the end of the 1990s. NSB's estimates of the foreign-born share of PhDs from the 2000 Census were 11 percentage points (41 percent) higher than NSF's estimates; NSB's estimates of the foreign-born share of MS scientists and engineers were 9 percentage points (45 percent) higher than NSF's estimates; and NSB's estimates of the foreign-born share of scientists and engineers were 5 percentage points (42 percent) higher than NSF's estimates. The reason for these huge discrepancies has to be the failure of SESTAT to count the new immigrants.

Table 5.1
Alternative Estimates of the Proportion of Scientists and Engineers Who Are Foreign-Born, by Degree

	PhD (%)	MS (%)	BS (%)
NSF 1999 estimate	27	20	12
NSB 2000 Census	38	29	17
CPS 1999 MORG	38	30	16
CPS 2000 MORG	42	28	16

SOURCES: NSB (2003); CPS data tabulated from the Merged Outgoing Rotation Group (MORG) files, available online at NBER.

Given the existence of the CPS, this problem can be solved in the future without a new Census-based survey. The CPS currently asks respondents their highest degree (rather than asking only years of schooling, as it did in the past) and their detailed three-digit occupation code. These data can be used to provide timely estimates of the foreign-born proportion in the scientific and engineering workforce (though not of the foreign-born proportion among persons with S&E degrees). Using the MORG CPS file at the National Bureau of Economic Research (NBER), I have estimated the percentage foreign-born among PhD, MS, and BS scientists and engineers for 1999 and 2000. The bottom two lines in Table 5.1 give my estimates of the foreign-born share of scientists and engineers in these years. The CPS-based estimates are close to the Census estimates and are far larger than the NSF estimates. Since the CPS is an ongoing regular survey, it can be used to update estimates of the foreign-born share of scientists and engineers annually.

While the CPS provides timely data on some aspects of the S&E workforce, it has too small a sample size to provide much detail on persons with specific doctorate specialties, postdocs, H1-B visas, and so on. In addition, the CPS does not currently ask about the field in which recipients earned their degrees, so it cannot measure the number of scientifically trained persons working outside the S&E occupations. There are, however, a large enough number of observations on bachelor's degree engineers to permit detailed analysis of that group. NSF should make greater use of the CPS to obtain data on engineers and scientists and should examine ways to combine CPS files over several years to estimate the job market situation of more narrowly defined groups.

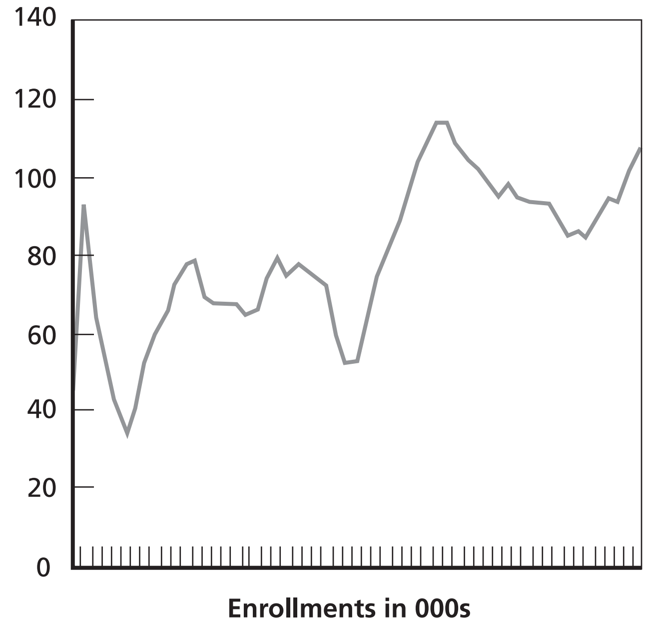
The other way to obtain more-timely data to illuminate issues in policy debate is to conduct "snap surveys" of S&E workers through email and the Internet. I discuss this further in my concluding recommendations on ways to improve our knowledge base.

Weakness of Data for Supply Behavior

Information on the number of students enrolled in the S&E curriculum and on the number of graduates, for which we have population counts, tells us nothing about the factors that determine the supply of young persons to science. If the numbers enrolled or graduating were relatively fixed, absolutely or relative to the number of young persons in some relevant choice set, we would not have to worry about labor supply behavior. But in fact, the numbers choosing S&E majors vary considerably, whether measured in absolute terms or relative to some base, such as all college and university enrollees or graduates, persons aged 18–24 or 25–29, and so forth. Figure 5.1 shows some of the variation in enrollments among bachelor's engineers, which evinces considerable oscillations of the "cobweb" type (Freeman, 1971, 1976).

There are two ways to study the determinants of supply behavior. The first methodology is to link the variation over time in numbers enrolling or graduating in different fields to measures of the potential incentives to enter the fields—salaries, ease of obtaining jobs, stipends, and the like—compared to the incentives to choose other careers, such as medicine, law, and business. Studies that have followed this route generally find positive relations (Freeman, 2003), indicating that students are sensitive to labor market conditions but lack information about where young people learn about the job market, the alternatives they con-

Figure 5.1
Oscillations in Enrollments in Engineering, Fiscal Years 1946–2001



RAND *CF194-5.1*

SOURCE: Freeman (2003).

sider when they make career decision, and the impact of their schooling experience on such decisions. These studies can make only the crudest assessment of the extent to which young persons look into the future as opposed to responding myopically to current market factors, and so on. In addition, the time series data on enrollments and degrees do not differentiate the high-ability students from others.

The second methodology is to survey students at the time they are making career decisions. Such a survey would determine what information students have, their perceptions of career opportunities in S&E, and the factors they believe led them to choose or reject S&E careers. Did they reject a scientific career because they can make more in other areas? Because the length of training or the riskiness of becoming a successful researcher is too high? Did they choose a scientific career because they had a particular experience in school, or because they won a fellowship? Without rich knowledge of the factors that enter into such decisions, it is difficult to imagine how the government could devise "strategies that attract high-ability students and professionals to S&E careers." **The single most important new data gathering and research that NSF or other groups concerned with the flow of young Americans into S&E could undertake would be a national survey of young persons, focused on what they know and are thinking about science careers during key decision periods.** Recognizing that what people say about their decisions at a point is unlikely to tell the whole story, this survey should have a longitudinal design, following the students as they proceed with further education, choose their careers, and enter the job market. It should sample all young persons with sufficient ability to go on in science or engineering, not just those who choose S&E careers. This would enable researchers to analyze the factors that lead some to select S&E careers and others to select other careers.

The United States has potentially valuable information about factors that affect the decisions of minorities and women regarding S&E careers. In the past 30–40 years, universities, foundations, and government agencies have developed diverse programs to attract highly able persons from underrepresented groups into S&E. There are summer programs; mentoring programs; special scholarships, such as NSF's Minority Graduate Fellowship (which operated from 1978 to 1998); Ford Foundation minority fellowships; NSF's Women in Engineering fellows; and so on. But information about the outcomes of these programs is scattered, making it difficult to judge what succeeded and what didn't, at what cost. Greater information on these programs would provide valuable insight into the career decisions of all young people regarding science, as well as about the effect of the policies on minorities or women.

Young Americans constitute only part of the U.S. supply of S&E workers. Two other groups are also critical to our nation's ability to conduct R&D: foreign-born students who study in the United States and foreign-trained scientists and engineers who immigrate to the United States. Foreign-born students make three critical supply decisions: to choose scientific and engineering careers, to study for those careers in the United States rather than in their home countries or at competing foreign sites, and to stay and work in the United States or to work elsewhere. We know little about the factors that influence these decisions and much less about policy interventions that might affect the decisions, such as immigration restrictions or the ease of obtaining green cards and citizenship. Some graduates from China or India wait until they get a green card before going home, to ensure that, if things go poorly in their home country, they can readily reenter the U.S. job market. A survey of persons who study S&E in the United States would lack the natural comparison group of highly able foreign-born students who chose other alternatives. But it would still be valuable to contrast the factors that foreign-born students say influenced their decision to study S&E in the United States to the factors that U.S. students say influenced them to enroll in S&E programs. Perhaps NSF could work with selected other countries, particularly those that send many students to U.S. programs, to study the decisions of the young to enter scientific careers on a comparative basis.

We also know little about what foreign-born graduates of U.S. universities do when they return to their home countries—whether they apply the knowledge they learned to set up businesses that compete with U.S. businesses or that complement U.S. businesses. A survey of the "alumni" of U.S. higher education who returned to their native lands would illuminate the factors that led them to leave the United States and the benefits and costs involved in their return migration. This would help the United States decide whether the country should try harder to retain these graduates or should welcome and encourage them to return home.

As indicated earlier, we have the least information about scientists and engineers who obtain their degrees overseas. This is a major weakness in our data system because foreign universities are an important source of supply now and are almost certain to become an increasing source of supply in the future. In the past 20 years, the number of PhDs graduating in S&E outside the United States has been increasing relative to the number graduating in the United States. In 1999, the United States produced approximately 26,000 S&E PhDs, compared to 39,000 in the European Union (EU) and 22,000 in Asia (NSF, 2002, appendix table 2-42). The U.S. share of global PhDs in S&E, which was on the order of 25 percent in

1999, will continue to fall in the future. What factors lead graduates of foreign universities to come to the United States? How important are

- salary differences between the United States and graduates' native countries?
- research facilities?
- the presence of outstanding scientists and engineers?
- the quality of American life, with its multiple opportunities and choices?

In addition, how many are recruited by American universities or firms, as opposed to taking the initiative to come themselves? Can the United States continue to attract top scientists and engineers from the EU, in the face of EU efforts to attract them?

Our current information base provides almost no insight into these questions.

Finally, it is critical to have more information about the elasticity of supply of the current S&E workforce to different fields, sectors, and projects. In the short and intermediate run, new research programs necessarily draw most of their workers from the existing stock of engineers and scientists and engineers, which in turn reduces the supply in other activities. From which activities does the nanotech initiative draw researchers? bioshield? defense R&D? How easy or hard is it for researchers to cross sectoral or field boundaries in response to new programs? One way to study this responsiveness is to compare the career rewards between the source sectors and disciplines from which the newly engaged researchers came to those for the new initiatives. A key factor in the flexibility or responsiveness of the existing workforce is likely to be the overlap in competencies between the new area of work and the older area. SESTAT provides information on fields of study and occupation but does not ask about the transferability of competencies—whether someone with training in plasma physics could readily learn the skills needed to operate in some nanotech activity, for instance. Data on what people can do and how quickly they could learn new skills should be more important than data on their fields of study years ago.

Another way to study responsiveness is to apply econometric analysis to data on numbers working and earnings in different sectors and projects when they experience policy-induced expansion or contraction. The most recent analysis of this type suggests that supply is relatively inelastic.

Weaknesses of Data on Demand and Production of Basic and Applied Knowledge

The types of work done by S&E workers have changed greatly in recent years. More S&E workers are employed in industry and academe; more work in large research centers; and more academic researchers are connected with commercial projects, initiated by faculty or by firms that establish links with universities. Surveys of scientists and engineers have not kept pace with these developments. We have no reliable national data on the growth of team science, much less on how this has changed the career activity of scientists and engineers. And we have very limited data on the scientific productivity, the allocation of credit on projects, and economic rewards in industry. To understand the production process by which S&E work produces new knowledge and the derived demand for S&E workers, we need information on the organizational mode under which scientists and engineers work in various sectors of the economy: the sizes of their work groups or laboratories; the allocation of decisionmaking and rewards in team activities; the different disciplines, skills, and activities of team members; the equipment used; and so on.

The SDR records the numbers of publications and patents reported by the respondents over two years. But it does not ask about the process that led to the publishable papers or patents, the numbers of coauthors or others who contributed to the project, or the other ways scientists and engineers contribute to the growth of knowledge and potential production of useful goods and services. This means that we have relatively less information on the work activities and productivity of scientists and engineers in industry, which is the growth area for employment. Adding a module about the way scientists and engineers work, focused on industry and team activities, to the SDR could greatly enhance our understanding of what is involved in science careers today and thus on the types of students we might expect to be attracted by them.

Finally, modern labor market analyses make extensive use of matched employee-employer data files, which contain information on the characteristics of workers and the characteristics and productivity of their firms. Matched files for scientists and engineers in industry would offer new insights into the link between the training and knowledge of the workers and the output of the firm and thus would help us understand the elusive transformation of education and ideas into products. If firms were also asked about the ease or difficulty of hiring researchers for projects and the existence of unfilled budgeted positions, these data could also help measure alleged shortages.

Ten Suggestions

Implicit in the preceding comments are suggestions for new initiatives in data gathering and analysis to fulfill NSB's imperative to have the federal government "substantially raise its investment in research that advances the state of knowledge on international S&E workforce dynamics." I have organized these comments into ten suggestions, some of which NSF, NIH, or other agencies could readily undertake and some of which would require major new investments in information. Table 5.2 provides a capsule summary of my ten suggestions.

1. **Use the CPS more.** NSF should use the CPS to analyze the pattern of change in the S&E job market and, in particular, to measure the number of foreign-trained scientists and engineers working in the United States. Information from the CPS should regularly be compared to information from SESTAT and its underlying surveys, and efforts should be made to understand any differences in the patterns. NSF should consider funding a special CPS supplement on scientists and engineers that could ask detailed questions about work activity, skills competencies, career patterns, and the like.
2. **Use Web surveys and conduct snap surveys on policy-relevant issues.** Agencies should use email and Web surveys to gather information on a timely basis at low cost. Conducting national surveys on the Internet has problems—the nonrepresentativeness of Web users and the lack of a well-determined population base from which to draw samples. These problems should be minimal for scientists and engineers, nearly all of whom should have emails. It should be possible to obtain addresses for well-defined populations of these persons. Graduates could provide emails on the SED. Recipients of grants, stipends, and other rewards should provide emails that would go into a govern-

Table 5.2
Ten Suggestions for Improving the Database and Analysis of the Science and Engineering Workforce

1. Use the CPS more.
2. Use Web surveys and conduct snap surveys on policy-relevant issues.
3. Gather data from Internet recruitment sites.
4. Survey young decisionmakers and follow their careers.
5. Create a clearinghouse of government and other interventions.
6. Survey working conditions of scientists and engineers.
7. Add competency questions to the SDR.
8. Developed a matched employer-employee data file from the Survey of Industrial R&D.
9. Increase ease of use and visibility of available data.
10. Hold annual or biannual research conferences on science workforce issues.

ment S&E email address bank which would be the base for low cost timely surveys. The key analytic question is whether response rates would be high enough to yield unbiased estimates. This would have to studied carefully, say by comparing answers on a low-response snap survey to those on a higher-response survey. The federal government should be able to get a high response rate from one important group of researchers: those getting federal research funding.

With an email database on scientist and engineering workers, the NSF would have the potential to conduct "snap surveys" on particular S&E workforce issues or to subcontract such surveys to the private sector. Such surveys would be infrequent and focused on specific issues, such as the situation of postdoctorate workers, or potential supply responses to new initiatives.

3. **Gather data from Internet recruitment sites.** Many graduate workers, particularly engineers in the information technology sector, make extensive use of Internet recruitment sites to search for jobs. Employers list job openings on these sites or on own their own Web sites. The information on the sites provides an opportunity to assess supply and demand conditions, to ascertain which skills are in great demand, and to discover what types of specialists have trouble finding work. The government should fund a study of the way the information on these sites can be used to create new indices of market conditions, comparable to the help-wanted indices that the Conference Board and other groups have used to assess labor market conditions in the economy more broadly.
4. **Surveys of young decisionmakers.** The federal government should fund new surveys on the three primary groups that provide scientists and engineers to the U.S. economy: highly able U.S. students, foreign-born students who come to the United States to study, and foreign-trained scientists and engineers who come to the United States to work. The survey of U.S. students should concentrate on those with high ability, should cover periods of key career decisions, and should follow their careers in a longitudinal design. The survey should include persons who choose S&E careers and those who do not. The survey of foreign-born graduates of U.S. programs should ask them the reasons

they chose to study in the United States and should follow them to see where they work after they graduate, including those who leave the country. The survey of foreign graduates who come to the United States should identify the factors that made the U.S. attractive to them compared to their home country and other potential areas of work.

5. **Create a clearinghouse of government and other interventions.** NSF should gather in one place information on the multitude of programs designed to attract underrepresented groups into the sciences and engineering. The data in this national clearinghouse should be analyzed with standard meta-statistical tools to see what generalizations, if any, can be drawn from successful programs and failed programs. The clearinghouse should provide information on federal supports available for career development, including summer internships for students, opportunities to work in federal laboratories, and potential government specialists visiting college and high school campuses.
6. **Survey working conditions of scientists and engineers.** To understand the change in the nature of scientific work from solo investigators to team science, NSF, NIH, and other concerned agencies should fund a regular survey of laboratories that would focus on workplace conditions: numbers of persons on teams, structure of rewards and incentives, use of equipment, and so on. This survey should sample working scientists and engineers, from graduate student research assistants to directors of laboratories. It could be used to benchmark questions on the SDR.
7. **Add competency questions to the SDR.** To assess the flexibility of supply across disciplines and sectors, NSF should develop questions for the SDR that ask about specific competencies, in addition to questions about training and degrees. By a competency question, I mean questions like these: "Could you today work on a project that involved XYZ technology?" "How much training would you need to be able to work on such a project?" In the case of information technology, the technologies might involve particular programming languages. In the case of biological sciences, one could ask about selected bioinformatics statistical analyses, or about operating particular pieces of equipment, and interpreting results.
8. **Develop a matched employer-employee data file from the Survey of Industrial R&D.** The technology for creating matched employer-employee data files is most highly advanced in Europe; in some of the smaller countries, such as Denmark, one can follow the movement of virtually every worker across firms. The U.S. Census developed a matched file with the 1990 Census and has since done much more work in this area. NSF should seek to match the R&D data on the Survey of Industrial R&D with more detailed information on employees at the sites and with matched data on output, productivity, and other information at the level of the firm.
9. **Increase ease of use and visibility of available data.** The CPS is the most widely used data set for analyzing labor market issues in the United States. One reason is that CPS files are readily accessible to researchers. The National Longitudinal Surveys funded by the Department of Labor are also widely used because they are readily accessible. Although NSF has made great improvements in data accessibility, it is still difficult to use some NSF data, such as the raw files of the SDR and SED. It is harder yet to use administrative data on the S&E workforce from the files of NSF and NIH, among other agencies. Making these data easier to use will increase the number of researchers working on them and thus increase the stock of social scientists and others working to advance

"the state of knowledge on international science and engineering workforce dynamics," per the goals of NSB.

10. **Hold annual or biannual research conferences on science workforce issues.** Regular conferences that bring together policymakers, groups conducting surveys, firms, employee associations, and researchers will help create a larger working community on science workforce issues and direct academic research toward policy-relevant issues.

References

Citro, C. F., and G. Kalton, Committee on National Statistics, Office of Scientific and Engineering Personnel, National Research Council, *Surveying the Nation's Scientists and Engineers: A Data System for the 1990s*, Washington, D.C.: National Academy Press, 1989.

Freeman, Richard B., *The Market for College-Trained Manpower: A Study in the Economics of Career Choice*, Cambridge, Mass.: Harvard University Press, 1971.

_____, *The Overeducated American*, San Diego, Calif.: Academic Press, 1976.

_____, "Rising Inequality in Science Careers," paper presented at AAAS Symposium, Is Science a Deteriorating Wage of Life: Workforce Trends, Denver, Colo., February 15, 2003.

National Research Council, *Forecasting Demand and Supply of Doctoral Scientists and Engineers: Report of a Workshop on Methodology*, Washington, D.C.: National Academy Press, 2000.

National Science Board, Committee on Education and Human Resources, Task Force on National Workforce Policies for Science and Engineering, *The Science and Engineering Workforce: Realizing America's Potential*, Washington D.C., NSB 03-69, 2003.

National Science Foundation, *Science and Engineering Indicators*, Washington, D.C., 2002.

CHAPTER SIX

What Data Do Labor Market Researchers Need? A Researcher's Perspective

Paula E. Stephan[1]

Introduction

My comments focus on data needed to inform our understanding of scientific, technical, engineering, and mathematics (STEM) labor markets as well as our understanding of the productivity of the STEM workforce. I see this as distinct from data needed to forecast future workforce requirements.

Indeed, I want to make it clear from the outset that I see forecasts of STEM labor needs as a slippery slope. I would encourage anyone who has not already done so to study the National Research Council (NRC) report, Forecasting Demand and Supply of Doctoral Scientists and Engineers (NRC, 2000).[2] The report notes that a number of exogenous variables that affect scientific labor markets, such as war, recession, changes in the federal budget, and alternations in immigration policy, cannot be predicted with any accuracy; the result is difficulty in predicting labor market outcomes with any accuracy. The report also emphasizes the importance of updating forecasts frequently, if and when forecasts are made.

The message with regard to the perils of forecasting has clearly not been fully absorbed. Despite warnings and past problems and embarrassments, the temptation persists to forecast, especially shortages, using simple models that are not updated. Michael Teitelbaum provided recent examples (see Chapter Two); another example comes in the prediction that we will need 2 million nanotech workers by 2015.[3]

Not forecasting is not equivalent to not providing information on labor markets. We can learn a considerable amount about conditions in STEM labor markets by observing labor market trends. What's happening to salary, and more generally to compensation, is a key metric. In looking at salary data, it is important to realize that rising salaries are not a "bad thing" but instead can be an important way of signaling opportunities for talented individuals in STEM fields.

---

[1] Paula E. Stephan is with Georgia State University. The author wishes to thank Grant Black and William Amis for their assistance and comments on an earlier draft.

[2] The report indicated that forecast error can also result when models are based on incorrect assumptions about "overall structure, included variables, lag structure, and error structure." It also pointed out that "data used for estimation may be flawed or aggregated at an inappropriate level" (p. 1).

[3] See, for example, NSF (2001).

Doctoral Workforce

Here I want to focus on four questions related to the doctoral workforce: (1) For whom do we need data? (2) What kinds of data do we need? (3) What is the time frame? (4) What facilitates research on the doctoral workforce?

For Whom?

A primary need we have is to know characteristics of individuals available for work in STEM markets in the United States. Our current data on doctoral scientists is drawn from the Survey of Doctorate Recipients (SDR) administered by Science Resources Statistics (SRS) of the National Science Foundation (NSF) and focuses primarily on those who received doctoral training in the United States. The sampling frame for the SDR is the Survey of Earned Doctorates (SED), a census complied by SRS of all research doctorates awarded in the United States. This sampling frame needs to be expanded to include individuals who received training outside the United States, especially given the large number of foreign PhDs who come to the United States for further training in a postdoctorate position. While steps have been taken in that direction, it would be important to have a census, similar to the SED, of those who enter the United States with doctoral training in hand.

Such a census would allow us to know considerably more about who has received training and or worked in the United States at the doctoral level. For example, to date there are considerable holes in our information concerning the countries of origin of those working or trained in the United States. Shortly after September 11, we used the SED to produce a list of the number of PhDs awarded in STEM from 1990 through 1999 to temporary residents from countries targeted for more rigorous screening (Stephan, Black, et al., 2002). The list, which was eventually published in Science as an extended letter, was instructive, indicating that 5,469 temporary residents from these countries had been awarded degrees in the United States during this period. This represented 8.8 percent of all degrees given during the period to temporary residents and 3.0 percent of all PhD degrees awarded. Five hundred and eighty-three of these were trained in what were deemed sensitive fields.[4] Instructive as this was, we were unable to enumerate the number of individuals who came from "targeted" countries for postdoctoral training. There was simply no systematic data available on the countries of origin of these scientists at the time we did the study.

What Kinds of Data?

As noted above, salary and compensation data are key to monitoring STEM labor markets. These provide a good indication of whether bottlenecks are emerging. Combined with placement data, they also give programs key information about the outcomes for their graduates. Salary data are most important at times when individuals make key transitions in their careers—specifically, at the time individuals move from graduate school to take full-time positions or postdoctorate positions and when they move from postdoctorate positions to jobs. While the SDR provides longitudinal data on a sample of doctoral scientists throughout their careers, it has scant evidence on salaries at these key transition times, especially as

---

[4] Sensitive fields include nuclear and organic chemistry; chemical and nuclear engineering; bacteriology, biochemistry, biotechnology research, microbiology, molecular biology, and neurosciences; and atomic, chemical, molecular, and nuclear physics.

individuals move from graduate school to their first positions. This is because individuals are not picked up in the SDR sampling frame until they have been out one or two years. The data, by necessity, are also "thin" and can provide little detail by field and by institution at the time of transition, because the SDR is an all-purpose survey that follows careers, not just those in transition.

The SED provides an opportunity to learn about the salaries individuals are offered at the time they make the transition from doctoral student to jobs or postdoctorate positions. The survey is administered at the time the individual receives the PhD and poses a series of questions on postgraduation plans. But the picture is incomplete on two counts. First, no salary data are collected. It would be straightforward to add a question concerning salary for those who indicate that they have definite plans. The picture is also incomplete in the sense that 30 percent of those responding indicate that they have no definite plans at the time the SED is completed. While a sample of these are followed in the SDR, key information concerning this 30 percent is missed at the time of graduation. A considerable amount could be learned by a follow-up questionnaire, ascertaining whether they now have plans, and if not, why not. This is precisely the kind of indicator that could tell us a good deal about the labor market for STEM workers.

The SED also provides no information on alternative employment offers that new doctorates receive. Yet alternatives are a powerful indicator. If there are none, it says a considerable amount about labor market conditions; if there are multiple offers, it provides insights concerning what STEM workers value. Scott Stern, for example, showed the power of such analysis by demonstrating econometrically that scientists going to industry take a reduction in salary to work in a firm that allows them to publish (Stern, 1999).

Comparable information about plans, salary, and alternative offers is needed concerning those who are leaving (or trying to leave) a postdoctoral position. It is encouraging to know that SRS has recently received funds to begin the design of a survey of postdoctorates. Given that leaving a postdoctorate position for a more permanent position has become a major transition point in a career and that problems associated with the transition have generated considerable concern, a census administered by the host university would be more appropriate than a sample, especially if it were administered at the time the individual left the institution.

We also need to know considerably more about STEM workers in industry. Industry in the United States now hires close to 40 percent of all STEM PhDs (Stephan, Sumell, et al., 2004). Despite the prevalence, we have little information about STEM workers in industry. We don't know where the jobs are located—San Jose, Portland, Chicago, Atlanta? What industries they are working in—semiconductor, information technology, pharmaceuticals, etc.? What do their jobs entail?

There are two rationales for such information. First, it informs students, faculty, and policymakers about the types of jobs available in industry. Given the prevalence of industrial hires, it is important to let students and their professors know more about jobs in industries. Too often graduate students are taught only about job possibilities in academe and only stumble into jobs in industry, considering them to be second-best options. Information on outcomes could help educate students regarding the prevalence and opportunities of jobs in industry.

The second rationale is to provide information concerning the economy. One headline in 2003 emphasized the important contribution that the service sector is playing in enhancing economic productivity.[5] Much of this results from the adoption of new technology in the service sector. Currently, however, we have no way of knowing which industries in service are attracting the largest number of new STEM doctorates. Data on PhDs working in industry could provide such insight.

More generally, industry human resources data provide increased understanding regarding innovation. There is increased awareness that much innovative activity occurs outside research and development labs. Wes Cohen and his coauthors found, for example, that the largest source of ideas for new innovations comes to industrial labs through interaction with customers (Cohen, Nelson, and Walch, 2002). While the sales force involved in this interaction need not have PhDs, a highly trained, technical workforce can be key. Others have pointed to the key role that mergers and acquisitions play in innovation and the necessity of having scientists and engineers involved in seeking out and evaluating such opportunities.

In short, we can enhance our understanding of innovation by learning where STEM doctorates are working in industry and what they are doing. This is not a new idea. A National Research Council Workshop, Using Human Resource Data to Track Innovation, argued for collecting such information more than three years ago (NRC, 2002).

Productivity

We also need reliable information on the productivity of the STEM workforce. This requires matching STEM databases with databases concerning productivity. Such a match helps our understanding of what fosters and facilitates productivity. This is an important issue, given that productivity of the STEM workforce is highly skewed.

For example, it is estimated that 6 percent of publishing scientists write 50 percent of all articles.[6] A similar, although even more skewed distribution is found among those patenting, whether they are working in industry or in academe. In recent years, the ability to match patent data with individual data has been greatly improved, thanks both to data from the U.S. Patent and Trademark Office and the National Bureau of Economic Research's Patent Citation Data File. There is need for better data, however, with regard to publications.

Currently, Thomson ISI® provides the only systematic data concerning publications and citations. These data have several serious drawbacks, however. One is expense. Another is the extreme difficulty that one has in identifying unique individuals from the data, given that, in many instances, only last name and first initials are coded and that addresses are not systematically linked to individuals. By way of example, among the names of the top ten most-cited physicists of the decade, only two are readily found to correspond to unique individuals. Six of the ten are common names that do not correspond to unique individuals who belong in the top ten; rather, the collected research of multiple physicists with the same last name and first initial lead the name to be among the top ten most-cited physicists. Two of the names require further research to determine whether they are unique.

---

[5] See, for example, Hilsenrath (2003).

[6] The highly skewed nature of productivity was first observed by Alfred Lotka in 1926 in a study of 19th century physics journals. The distribution that Lotka found showed that approximately 6 percent of publishing scientists produced half of all papers. Lotka's "law" has since been found to fit data from several different disciplines and varying periods of time (Price, 1986).

Timely Release of Data

To be informative, data must be released in a timely manner. Especially important is the early release of key indicators, such as salary information, the number of offers individuals receive (if the above recommendations were to be adopted), and the sector of employment. Agencies should not shy away from such an early release, within, for example, three months of collection, given that the numbers can always be updated and that there is a tradition of updating early release data, such as productivity measures, in the United States.

Access Is Important

Creative work regarding STEM labor markets often comes from external researchers. In evaluating proposals from external researchers, federal statistical units should focus on the benefits to be gained by such research—not only on the risks associated with the use of the data by external researchers. Little creative research can be carried out if statistical agencies obsess about worst-case scenarios. Statistical units need to remember that faculty consistently work in a culture that respects confidentiality. They also need to appreciate that faculty researchers face different incentives from those of government employees. To wit, the ability to publish in a timely fashion is crucial to faculty. Yet involvement with a federal statistical agency adds considerable time to the research process. One must first negotiate for a license. This can take up to six months. Agency review of research is required, taking four to six months. And then it is possible, as has been known to occur, that the agency changes its mind about what is permissible and demands extensive revisions after the fact that add yet another six months or more to the research process. Such problems, and such timelines, send clear signals to the research community that involvement with a federal statistical agency is risky business.

Summary of Needs

To summarize, we need enhanced data, especially with regard to salary, alternatives, and industrial employment, especially for individuals at key career transitions, and we need this data for those trained outside as well as inside the United States. The timely release of these data is important if they are to inform our understanding of labor markets and provide input for those making career decisions. Moreover, we need good data on productivity, especially publishing productivity, that can be matched with microdatabases. Access to the data by external researchers is important. Such access requires agencies to focus on benefits as well as risks and appreciate that external researchers face different incentives than do government employees.

References

Cohen, W., R. Nelson, and J. Walch, "Links and Impacts: The Influence of Public Research on Industrial R&D," *Management Science*, Vol. 48, No. 1, 2002, pp. 1–23.

Hilsenrath, J. E., "Behind Surging Productivity: The Service Sector Delivers," *Wall Street Journal*, November 7, 2003, pp. 1 and A8.

Lotka, Alfred, "The Frequency Distribution of Scientific Productivity," *Journal of Washington Academy of Science*, Vol. 16, No. 12, 1926, pp. 317–323.

Mowery, D. C., "America's Industrial Resurgence: How Strong, How Durable?" *Issues in Science and Technology*, Vol. 15, No. 3, pp. 41–48.

National Research Council, *Forecasting Demand and Supply of Doctoral Scientists and Engineers: Report of a Workshop on Methodology*, Washington D.C.: National Academy Press, 2000.

_____, *Using Human Resource Data to Track Innovation: A Summary of a Workshop*, Washington, D.C.: National Academy Press, 2002.

National Science Foundation, Societal Implications of Nanoscience and Nanotechnology, March 2001. Online at http://www.wtec.org/loyola/nano/societalimpact/nanosi.pdf (as of June 3, 2004).

Price, Derek J. De Solla, *Little Science, Big Science ... and Beyond*, New York: Columbia University Press, 1986.

Stephan, P. E., G. C. Black, J. D. Adams, and S. G. Levin, "Survey of Foreign Recipients of U.S. Ph.D.s," *Science*, Vol. 295, March 22, 2002, pp. 2211–2212.

Stephan, P. E., A. J. Sumell, G. C. Black, and J. D. Adams, "Doctoral Education and Economic Development: The Flow of New Ph.D.s to Industry," *Economic Development Quarterly*, Vol. 18, No. 2, May 2004, pp. 151–167.

Stern, Scott, "Do Scientists Pay to Be Scientists?" Cambridge, Mass.: National Bureau of Economic Research, working paper 7410, October 1999.

CHAPTER SEVEN

What Data Do Science, Technology, Engineering and Mathematics (STEM) Agency Policymakers Need?

Judith A. Ramaley[1]

To answer the question posed in the title, it is important first to describe the mission and goals of the Education and Human Resources Directorate (EHR) at the National Science Foundation (NSF) and the areas of emphasis articulated in the cross-directorate priority area called Workforce for the 21st Century, which will shape how NSF addresses scientific, technical, engineering, and mathematics (STEM) education and workforce development. This will set the stage for defining what we need to know in order to decide how to advance this agenda.

As in all federal policy, there are several key steps that must be followed in order to understand how an agenda is set, how strategies are identified, and how the selected strategies are implemented and evaluated. Although there are many models of the policy process, this paper will employ Randall Ripley's approach to describing the stages of the policy process, which provides a guide without offering an explanation of how policy comes about (Ripley, 1997, pp. 3–16). The elements that will be developed in this paper are (1) agenda setting, (2) formulation and selection of goals, (3) program implementation, (4) evaluation and assessment of impact of programs, and (5) decisions about the future of policy and programs.

Phase 1 Agenda Setting: The Overall Functions of the National Science Foundation

According to 42 U.S.C. § 1861, as established by Congress in 1950, the functions of NSF include, among other things, to

> initiate and support basic scientific research and programs to strengthen scientific research potential and science education programs at all levels in the mathematical, physical, medical, biological, social and other sciences

as well as in engineering; to award scholarships and graduate fellowships for study and research in the sciences or in engineering; to provide a central clearinghouse for the collection, interpretation, and analysis of data on scientific and engineering resources; and to provide a source of information for policy formulation by other agencies of the federal gov-

[1] Judith A. Ramaley is Assistant Director, Education and Human Resources Directorate at the National Science Foundation in Arlington, Virginia. She also holds a presidential professorship in biomedical sciences at the University of Maine-Orono and is a Fellow of the Margaret Chase Smith Center for Public Policy. The opinions expressed in this paper do not necessarily represent the official views of the National Science Foundation.

ernment. To accomplish its data-gathering functions, NSF prepares a report on S&E indicators every two years.

Phase 2 Formulation and Selection of Goals: The Goals of the National Science Foundation

In its Government Performance Results Act of 1993 (GPRA) Strategic Plan for fiscal years 2003–2008, NSF established four overall goals. Three of these are programmatic goals that address workforce issues in both direct and indirect ways:

- ***People Goal:*** A diverse, competitive, and globally engaged U.S. workforce of scientists, engineers, technologists, and well-prepared citizens.
- ***Ideas Goal:*** Discovery across the frontier of science and engineering, connected to learning, innovation, and service to society.
- ***Tools Goal:*** Broadly accessible state-of-the-art S&E facilities, tools, and other infrastructure that enable discovery, learning, and innovation.

Phase 3A Program Implementation: A Capsule Portrait of the Education and Human Resources Directorate at the National Science Foundation

EHR's mission

> is to achieve excellence in U.S. science, technology, engineering and mathematics (STEM) education at all levels and in all settings (both formal and informal) in order to support the development of a diverse and well-prepared workforce of scientists, technicians, engineers, mathematicians and educators and a well-informed citizenry that have access to the ideas and tools of science and engineering. The purpose of these activities is to enhance the quality of life of all citizens and the health, prosperity, welfare and security of the nation.

The following are the goals for EHR's portfolio:

1. to prepare the next generation of STEM professionals and attract more Americans to STEM careers
2. to increase the technological and scientific literacy of all Americans so that they can exercise responsible citizenship in an increasingly technological society and acquire knowledge of science, mathematics, and technology that is appropriate to the development of workforce skills and life-long career opportunities
3. to broaden participation (diversity) and achievement in STEM
4. to attend to critical workforce needs requiring significant math and science skills and knowledge, both by attracting new people to these STEM careers and by support for the development and retooling of the current STEM workforce.

EHR has the following capacity-building strategies:

1. identifying effective ways to prepare and support teachers and faculty who can inspire and challenge students in the STEM disciplines and to provide them with effective materials and strategies to promote and assess learning

2. investing in research in the science of learning, facilitating the translation of research into practice, and creating supportive learning environments and STEM pathways by developing models of reform or systemic change at both institutional and multi-institutional levels through networking, partnerships, alliances, and collaborations
3. ensuring that the STEM community is broadly representative of the nation's individuals, geographic regions, types of institutions, and STEM disciplines
4. identifying effective ways (formal and informal) to address the STEM knowledge requirements of adults so that they can be productive members of the workforce and informed and active citizens.

The EHR portfolio has five specific elements:

- supporting individuals to enhance the capabilities of those preparing for STEM careers (including K–12 teachers) and those currently in the workforce and to attract a diverse group of people to become scientists, engineers, technologists, mathematicians, and educators
- developing ideas, tools, and communication strategies that will enhance the quality of STEM education; inform all citizens about science, technology, and mathematics; and enhance public understanding of current research and its implications
- utilizing a solid body of evidence created by an analysis of the existing research and examination of practice to guide further development of the NSF portfolio, as well as to support the diffusion and implementation of strategies that promote educational excellence
- supporting the development of links among networks of researchers and practitioners that help create the capacity to introduce and sustain change to enhance the quality of STEM education, improve the relationship between research and practice, and broaden participation in STEM careers
- developing institutional environments and enhancing system-level interactions and partnerships that support change and the introduction of strategies that will enhance the preparation and development of the STEM workforce—creating the infrastructure to support collaboration across disciplines and across institutions that promote excellence in STEM education and facilitate the application of innovation to economic development.

The core programs in the directorate are directed toward individuals, institutions, and collaborations. All are infused with a research mindset and are operated within a culture of evidence:

- The Math Science Partnerships: comprehensive and targeted awards; Research, Evaluation, and Technical Assistance; summer institutes for teachers (in the planning stage)
- Teacher Professional Continuum
- Informal Science Education (public understanding of science)
- Centers for Learning and Teaching, both K–12 and higher education (preparing the next generation of faculty and enhancing the preparation of K–12 teachers)
- The Experimental Program to Stimulate Competitive Research (EPSCoR)
- Louis-Stokes Alliances for Minority Participation (LSAMP) and the Alliance for Graduate Education and the Professoriate (AGEP) (broadening participation)

- Science and Technology Expansion Program and Course, Curriculum, and Laboratory Improvement (enhancing the undergraduate experience)
- Support for graduate students (attracting more U.S citizens and permanent residents to advanced study and exploring fresh approaches to the graduate experience): Graduate Research Fellowships (GRFs), Integrative Graduate Education and Research Traineeship program (IGERT), Graduate Teaching Fellows in K–12 Education program (GK–12)
- Research on learning and teaching: Research on Learning and Education (ROLE), Interagency Education Research Initiative (IERI).

Phase 3B Program Implementation: Workforce for the 21st Century Priority Area

In addition to the programs in the EHR portfolio and the programs that address workforce investments in the other directorates and offices of the foundation, the agency proposes to establish a cross-directorate priority area designed to identify and support effective strategies within the context of the disciplines and emerging areas of science and engineering supported by NSF that

- prepare scientists, mathematicians, engineers, technologists, and educators capable of meeting the challenges of the 21st century
- attract more U.S. students to science and engineering fields
- broaden participation in science and engineering fields.

Phase 4 Evaluation and Assessment of Impact of Programs and Phase 5: Decisions About the Future of Policy and Programs

The education and workforce development programs designed and implemented by NSF focus on one of three levels:

- support for individual advanced study
- support for institutional capacity-building
- support for collaborations, partnerships, and alliances that create pathways to excellent STEM education.

NSF utilizes a number of approaches to assessment, including consultation with Directorate advisory committees, evaluation of programs every three years by committees of visitors, the collection and consideration of research findings and project outcomes in the GPRA process, and larger-scale portfolio analysis.[2] All these methods examine the outcomes of individual awards or clusters of related projects within a program.

Our greatest challenge is to identify methods that will permit us to study the longer-term impact of programs that address large-scale interventions across time and across institu-

[2] The methodology for portfolio analysis is being developed in EHR and will be applied to the thematic interpretation of the quality, relevance, and productivity of investments in key programmatic areas, such as technology and education and mathematics education.

tions and to provide a solid quantitative basis for determining whether we should increase our investment in STEM education and in research on education and learning and, if so, what the goals of that investment ought to be. In doing so, we are less interested in whether there is a shortage or surplus of science and engineering personnel but rather in understanding the broader dynamics that will affect adult competencies and capabilities in an increasingly technological and knowledge-based economy. We are also deeply concerned about the significant gaps that still exist in participation and achievement in STEM fields. These gaps limit opportunity and threaten social cohesion in an increasingly diverse nation.

With respect to our first challenge, we have been working with our colleagues in the Social, Behavioral and Economic Sciences Directorate to address the following research issues:

- translating well-researched models and research-tested practices into approaches that are appropriate for a particular context, its assets, its culture, the nature of the local community, and the local economy (*dissemination*)
- moving beyond promising pilots or small-scale projects to large-scale change (*scale-up*)
- developing effective strategies that engage scientists and engineers from all disciplines in the work of reform (*partnerships with the disciplines*)
- creating the capacity to sustain the work over time (*sustainability*)
- creating a credible body of evidence upon which educational innovation and policy development can be based (*a culture of evidence*).

With respect to our other challenge of offering a well-researched rationale to direct our investment in education and workforce development, we need (1) *data* that will allow us to determine whether the following assertions are true and, if so, (2) *timely* information that will allow us to monitor changes in the national experience in these areas. Each of these assertions is followed by an example of one or more questions for which accurate and timely data are needed:

1. **We are overly dependent upon attracting well-prepared international students to fill many of our most-skilled science and engineering positions and to study in our graduate programs in these fields.** Only the three graduate fellowship programs in NSF's Division of Graduate Education require U.S. citizenship or permanent residence. We must expand the number of fellowships available in GK–12, IGERT, and the GRFs. We have increased the stipend level but need to add an appropriate increase in the provisions for educational costs that have remained constant for several years while cost of education has continued to rise. The demand for these programs greatly exceed our current funding level, and we could easily double our support without running out of highly qualified applicants.

   Related Data Questions: Can we trace the changing global movement of students as well as individuals with advanced education in STEM fields? Where do they study? Where do they then seek employment?
2. **The demography of this country is shifting dramatically toward a much more-diverse nation.** However, we have significant gaps in the participation and achievement of women and minorities in STEM education and in many STEM professions. As the opportunities for STEM professionals have increased, we have not seen a comparable

increase in enrollments. In fact, in some areas, such as mathematics, the enrollments are dropping. This problem will be exacerbated by the changing composition of our society. The participation of minorities and women in STEM professions is significantly less than that enjoyed by white males, especially in such fields as computer science, engineering, geosciences, mathematics, and the physical sciences. We have promising results from our programs for minority participation, such as LSAMP, about how to attract and graduate students of color in STEM fields. We must expand these programs and link them up effectively to graduate study and entry into the profession. Our program that supports this pathway is AGEP. It also is badly oversubscribed, and we need to invest both in financial support for students who need help in managing the financial burdens of further study and in assisting the sending and receiving institutions in providing effective educational and research opportunities for these students. The students that we most need to attract to STEM fields study predominantly in community colleges, minority-serving institutions, or regional public institutions that often lack sufficient capacity to provide effective educational programs and research experiences that can attract and sustain the interest of students in STEM. We must put more resources into developing the capacities of these kinds of institutions in science, technology, and engineering.

Related Data Questions: Are opportunities actually increasing? Where do students from underrepresented groups study STEM fields, and what is the productivity and quality of the choices available to them? What factors influence the decisions of individuals to pursue study in STEM and to enter these fields?

3. **Too few high school graduates are prepared to pursue the study of STEM fields at the postsecondary level.** The statistics here are alarming both for the future STEM workforce and for the general competency of the U.S. workforce. As many corporate leaders have said, adequate literacy skills used to be enough to make a living wage in our economy. Now, a facility with mathematics, technology, and science is also essential for productivity in the workplace, along with much-higher literacy skills. These abilities are required not only for those who enter STEM fields but also for the general workforce and for responsible citizenship.

   Related Data Questions: What competencies are now sought by employers? Are employers having difficulty finding well-prepared applicants and, if so, in what sectors or disciplines or fields? In cases where there appears to be high demand, how are employers responding? Are they sending jobs overseas? If so, which ones and in what quantity? Are corporations doing so to position themselves in emerging markets or because of a scarcity of skilled workers in the United States, or both?

4. **Too few K–12 teachers of science and math have adequate preparation and professional support to provide excellent instruction,** and the anticipated rates of retirement and the unacceptable loss of qualified teachers from the field (average turnover rate is about 5 years) prevent us from making much of a dent on the competencies of the teacher corps. We have introduced the concept of Centers for Learning and Teaching both at the K–12 level and in higher education to address the preparation of the next generation of teachers, as well as the teachers of those teachers, and to explore ways to enhance the undergraduate experience of all STEM majors and the preparation of future

faculty. We need to increase the network of centers significantly to create the capacity to address these significant workforce and education problems.

Related Data Questions: How can we measure "teacher quality" and monitor the development of competency in the teaching profession?

5. **The resources devoted to research on science and math learning and on educational improvement are woefully insufficient.** In the face of a growing national need for research-based solutions to our educational challenges, we devote less than one-half of one percent of the resources invested in public education (K–20) on research on education. In addition, we know very little about how adults learn, develop expertise, and remain skilled in a rapidly changing technology-driven society. We must at least aim for investing 2 to 3 percent of our resources in discovery, innovation, and application of research to educational practice.

   Related Data Questions: Do we have accurate data on how much is spent on education research and on research on learning? How can we follow the potential impact of this work on educational practices and student outcomes?

6. **We do not know much about why people choose careers in STEM or the pathways they travel to prepare themselves and then to remain skilled and competitive.** We also do not have the tools to model the STEM workforce or to predict changes in our overall capacity. We must add significant resources to our research support for developing these models and tools as a guide to federal and state education and workforce policy. We have practically no efforts in this area today. A significant investment in the Workforce for the 21st Century Priority Area and the STEM Talent Expansion Program will help us move more quickly toward an ability to understand and then begin to support a more-effective workforce policy package and investments in the capacity to educate and support STEM professionals.

   Related Data Questions: How can we better model the dynamics of the STEM workforce, including what kinds of jobs are occupied by STEM professionals and how the workforce adjusts to changes in both supply and demand? What data can we collect to help us assess both supply and demand?

7. **There is a growing demand for well-prepared technicians and support personnel in STEM fields.** We have developed an impressive program in advanced technological education in the past decade. We need to put more resources into preparing this part of the STEM workforce. There is also much interest in providing greater access to professional science master's programs for people who will manage laboratory facilities, support public policy, guide economic development strategies, etc.

   Related Data Questions: What data can be gathered that will allow us to assess the assertion that there are significant opportunities for individuals who hold a professional master's degree or a technical degree?

8. **Public understanding of science has never been more important.** Most of funding for NSF's efforts to promote public understanding of science and to link informal science education facilities and programs with the formal educational system remains flat. In addition, we need to develop new methodologies to measure or estimate public understanding of science and technology and the effectiveness of our efforts to support access to informal science education. As investigators increase the integration of research and education in the Research and Related Activities Directorate's portfolios, they are

exploring various kinds of outreach, especially as part of the programming of large centers, such as the Science and Technology Centers and Engineering Research Centers. However, our lack of knowledge about how adults learn, or how people use science museums and other informal programming and what they learn from these experiences, will continue to limit both our effectiveness and our ability to measure the value of these efforts. We must put more resources into research and development.

Related Data Questions: What can we do to measure adult "understanding" of science and the efficacy of various forms of outreach by scientists and engineers, as well as the impact of informal science education programs?

9. **Many of our expectations for improvements in STEM education require transformational changes in our nation's educational institutions,** both K–12 and higher education, as well as in informal science education, yet we know very little about the underlying processes of organizational change and the leadership of change or how to facilitate it. For a number of years, NSF has put small amounts of funding toward these issues. However, the awards have predominantly supported studies of change in industry, not in education, although there have been a few education awards from time to time.

   Related Data Questions: What data could we collect that monitors institutional change? We need to build the capacity of the research community that could conduct such studies and establish methods for measuring changes in institutional performance and function.

10. **The introduction of ideas developed through research, as well as choices based on research findings, into both educational policy and practice is extremely difficult.** According to Robert Rich (University of Illinois at Urbana-Champaign), public policymaking can be thought of as a process of public problem-solving during which elected representatives or educational leaders formulate and implement "solutions" to publicly defined problems. These problems are identified not by scholars but by a process of advocacy and the resolution of often conflicting interests. This is more commonly a bargaining process than a learning process. This contrasts sharply with the mechanism by which scientific ideas and findings are validated. The validity of new ideas and the significance of experimental findings will be judged by scholars according to their significance in terms of other ideas and other experiments within a body of literature and experience accumulated through the use of research methodologies. The model of trustworthiness generated by peer review is very different from the problem-solving process that takes place in the political arena. It is also different from the criteria implicitly applied by educational leaders as they make critical choices in practice. As policymakers and practitioners look at the ideas and findings generated by researchers, they ask

    - Is the work relevant to my sphere of responsibility? (the relevance test)
    - Is the work trustworthy? (the truth test)
    - Does the work provide clear guidance and direction in my circumstances? (the utility test)

    In sum, a policymaker or practitioner will ask, "Am I sure that the work is sound and that it has been interpreted fairly and not according to an ideological position? If I rely upon this, will I have a defensible position if and when my decisions are challenged?"

Related Data Questions: We need to develop methods that will allow the effective communication to policymakers and practitioners of ideas and funding generated by research. What data will we need and how can we ensure that these data are convincing?

Reference

Ripley, Randall B., "The Nature of the Policy Process," in Lester F. Goodchild, Cheryl D. Lovell, Edward R. Hines, and Judith I. Gill, eds., *Public Policy and Higher Education*, ASHE Reader Series, Pearson Custom Publishing, 1997.

CHAPTER EIGHT

What Data Do STEM Agency Policymakers Need? Workforce Planning for the Future: The NASA Perspective

Patrick Simpkins[1]

Data on scientists, technologists, engineers, and mathematicians are essential to mission success at the NASA. To explain the types, sources, and use of the data needed by policymakers, this paper first describes the role and influence of data in the development of NASA's strategic human capital management architecture; then discusses specific types and sources of data; addresses the uses of the data in human capital management decisionmaking; and, finally, offers suggestions for additional data structure and design that may be considered by the Office of Science and Technology Policy.

Data as Part of an Integrated Picture

Public- and private-organization policymakers require various types of data from a myriad of sources to be converted to information. These data must be collected throughout an organization's workforce planning and analysis process. NASA has developed an integrated workforce planning and analysis system that allows managers to gather data, convert the data into information, and make decisions based on that information. NASA desires and acquires information relevant to the past workforce, existing workforce, and the projected workforce. This information enables the agency to understand workforce drivers and thus manage its human capital more effectively.

Following a decade of downsizing, initial data indicated that NASA was facing skill imbalances, lack of depth in critical competencies, and potential attrition surges in the coming years. At the same time, there was evidence that NASA was facing greater competition in attracting the talent needed to ensure mission success for the future. The health of the workforce is a key concern of senior management, since a robust, agile workforce is essential to mission success. NASA must be able to attract and retain the talented men and women who will help mold the future of our nation's aeronautics and space programs. The agency must continue to have the scientific and technical expertise necessary to preserve our role as

[1] Patrick Simpkins is a systems engineer and human resources manager in the Information Technology Systems Integration unit of NASA's Management Systems Division, Washington, D.C. A great deal of the content of this paper is from previous efforts on the part of the National Aeronautics and Space Administration to develop and implement an integrated workforce planning and analysis system. Individuals such as Carolyn Davis and Carol Saric have developed content for parts of the agency's workforce legislation implementation planning and strategic plans, respectively. Further, the agency strategic planning teams and strategic human capital implementation planning teams are all contributors, directly or indirectly, to the development of the content for this paper. I would also like to thank Craig Conlin, the architect of the workforce planning and analysis web tools, and Chris Carlson, the designer and developer of the Competency Management System (CMS) for their direct, and indirect, contribution to the content of this work.

the world's leader in aeronautics, space, biological, physical and earth science, and emerging technology research.

These circumstances highlighted the need to collect more-comprehensive and accurate workforce data, analyze it in a more systematic and integrated manner, and use the analyses to develop strategies and initiatives to combat adverse workforce trends.

Existing data confirmed that certain workforce trends needed to be reversed. Within the science and engineering (S&E) workforce, the over-60 population outnumbers the under-30 population by nearly 3 to 1, and 25 percent of that workforce will be eligible to retire within five years. The potential departure of these individuals could deprive NASA of a wealth of knowledge, experience, and leadership essential to achieving its goals and objectives. As an agency with a highly technical mission, ensuring knowledge transfer from the senior workers to a new generation of employees is critical. The agency must be able to provide incentives to retain experienced employees as needed to ensure continuity on critical projects while also mentoring new talent.

The data also confirmed the lack of diversity in the workforce. Approximately 80 percent of NASA's scientific, technical, engineering, and mathematics (STEM) workforce is white and male.

During the downsizing of the 1990s, when the agency was not hiring external candidates in significant numbers, workforce strategies did not rely heavily upon analyses of trends in the U.S. labor market or upon shifts in disciplines being studied at institutions of higher education. Renewed efforts to revitalize the workforce, however, do require close attention to data of this kind. Examination of data and analyses from external sources (such as data from the Bureau of Labor Statistics, the Education Department, and the National Science Foundation) indicated that the adverse trends within NASA's workforce would be exacerbated by the external environment.

Perhaps the most significant trend involves the S&E pipeline, which has been shrinking over the past decade as undergraduate and graduate enrollment declines in disciplines critical to NASA's mission (Grossman, 2002; National Science Foundation, 2002; O'Keefe, 2003). In addition, while the number of women and minorities in these disciplines is increasing, they are still underrepresented in the undergraduate and graduate S&E student pool.

At the same that the S&E pipeline is shrinking, the demand for technical talent is increasing. Data reveal that the job market in S&E fields is projected to increase over the next decade. This projection is based in part on the anticipated retirement of baby boomers employed in the private sector, which creates new job opportunities. In addition, the demand for technical expertise is no longer confined to the technical industries; STEM graduates now are sought after in traditionally nontechnical businesses (such as banking or entertainment) not previously considered primary choices for technical graduates. Competition from the private sector for STEM talent has intensified with these competitors offering better tools and incentives to attract the high quality, diverse applicants.

In response to this information, NASA has taken aggressive action to address the workforce challenges it faces. NASA senior management has emphasized that the agency must have an integrated, systematic, agencywide approach to human capital management:

> NASA will implement an integrated Agency-wide approach to human capital management. This approach will attract and maintain a workforce that is representative

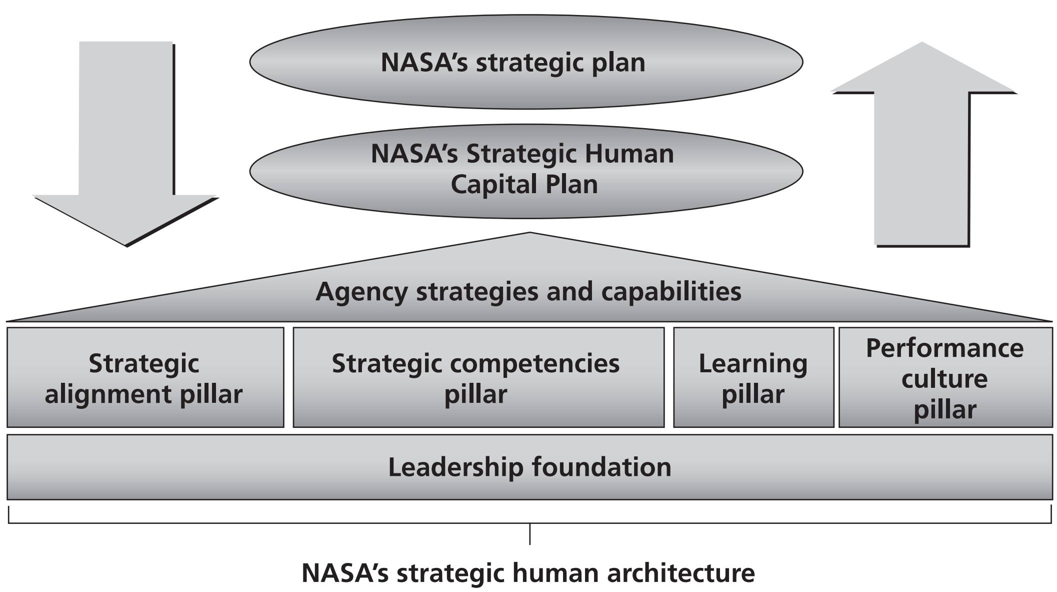
of the Nation's diversity and includes competencies that NASA needs to deliver sustained levels of high performance that the Agency's challenging mission requires. (NASA 2003 Strategic Plan)

The Strategic Human Capital Plan developed by the agency provides the framework to align human capital strategies and initiatives more effectively to NASA's vision and mission. The architecture is built on a foundation of strong leadership and supports the agency's mission and vision through the understanding and management of the "pillars" of the architecture: strategic alignment, strategic competencies, learning, and a performance culture (Figure 8.1).

The plan also lays out a clear set of initiatives designed to strengthen the leadership foundation and associated pillars. The success of each initiative is highly dependent upon the collection and analysis of key data and information. Among the initiatives NASA has undertaken, consistent with the goals of that plan are

- development of an agencywide, Web-based workforce planning and analysis capability, along with a competency management system to identify skills gaps in mission-critical areas, to ensure that the workforce is aligned to the agency mission
- improvement of the linkage between colleges and universities, NASA education programs, and projected workforce needs
- submission of legislative proposals to provide new authorities and tools to enable NASA to compete successfully with the private sector for a world-class workforce, reshape the workforce to address skills imbalances, leverage outside expertise, encourage students to pursue STEM careers, and build a workforce representative of the nation's diversity.

Figure 8.1
NASA Strategic Human Capital Architecture



RAND *CF194-8.1*

Types and Sources of Data for the Workforce of Today and Tomorrow

Assessing Workforce Competencies

Critical to the agency's ability to strategically manage its human capital is NASA's enhanced workforce planning and analysis capability. These tools include a data analysis engine that provides views of the total agency workforce, as well as of selected subsets for multiple years, and a forecasting engine that projects attrition for up to five years for any segment of the workforce. One of the most important components of this enhanced capability is the CMS.

In developing the CMS, NASA first ensured that there was a common set of terms and definitions across the agency that would be used to describe and identify competencies. Drawing on the agency's strategic, enterprise, and center plans, and other documentation and studies to ensure alignment with NASA's mission, goals, and objectives, a comprehensive Organizational Capabilities Dictionary was developed (NASA, 2003a). This dictionary identifies key knowledge areas critical to supporting enterprise and center strategies and achieving the agency mission. The next stage of analysis went further, resulting in the development of NASA's Workforce Competency Dictionary. It drew from components inherent in the organizational capability definitions and center workforce planning documents. It comprises critical competencies that reflect the knowledge or "know-how" needed to achieve NASA mission objectives. As appropriate, it will be reviewed to ensure that it is up to date and can continue to be utilized in gathering knowledge and experience data about the NASA civil service workforce.

NASA then applied attrition models to the competencies in the agency workforce inventory. Using employee demographic and retirement eligibility data, as well as turnover trends based on data from the workforce analytical tools, competency areas were identified in which there was potential for higher attrition than the agency average and thus would require greater focus in the development of the pipeline and near-term recruitment activities of the agency (Table 8.1, at the end of this chapter). Once competency data and workforce data were collected, further analysis and trending could be applied to the information so that informed decisions could be made to ensure that the current and future workforce has the competencies NASA needs.

The Future Workforce Pipeline Issues

To ensure a workforce with the right competencies needed in the future, NASA collects a range of data associated with the nation's pipeline of STEM talent. This includes data available from the Department of Education, Department of Labor, National Science Foundation, and other education and professional organizations. The data elements include demographic information associated with race, ethnicity, gender, and education.

Education data (numbers and fields of studies of graduates) related to STEM specialty areas are very important to NASA. The information is used to understand what academic disciplines are being studied now for near-term recruitment needs. More importantly, perhaps, it can be used to address future competency needs by influencing the disciplines students elect to pursue and the development of academic curricula through the use of grants, fellowships, scholarships, and student programs. For instance, nanotechnology applications are potentially far-reaching in many of NASA's missions; therefore, the support of this emerging area of study is beneficial to the agency.

Diversity of ideas and approaches is also important to NASA. It is essential to maintaining a robust environment for scientific inquiry and engineering excellence and ensuring an inclusive environment for the rich expression and exchange of ideas.

Another data set for workforce planning and analysis that merits greater attention than it has received to date involves economic trends—particularly with respect to emerging and declining industries and their impact on the labor market. For example, the state of the aerospace industry has been perceived as being in decline, which may influence STEM students to pursue studies other than aerospace engineering. Obviously, this is an issue of concern to NASA and other organizations that require competencies in that discipline.

Other study and survey data can shed light on the factors that influence an employee's career choices—such factors as the challenge of the work itself, compensation, work environment, work-life balance, and developmental opportunities. These are also of considerable importance to an agency's efforts to attract and retain a quality workforce.

Peter Drucker, as cited in Green (2000), proposed that the dominant factor in the 1990s and into the new millennium will be demographics. Information on the demographics of the existing, as well as future, workforce is essential to strategic workforce planning and analysis. The typical demographics of age, gender, race, ethnicity, and education are all important but should be augmented by trend data, such as

- typical postgraduate trends—where graduates are going and what they are doing
- attrition and retirement trends in addition to eligibility—who really leaves, and when
- levels of detail on both degrees and their content—classes taken
- knowledge application history and experience—where and how they used their knowledge
- graduation data—
  - at the degree program component level—what makes up the degree
  - areas of specialization—a degree in materials may be a specialty in polymers
  - areas of applied experience.

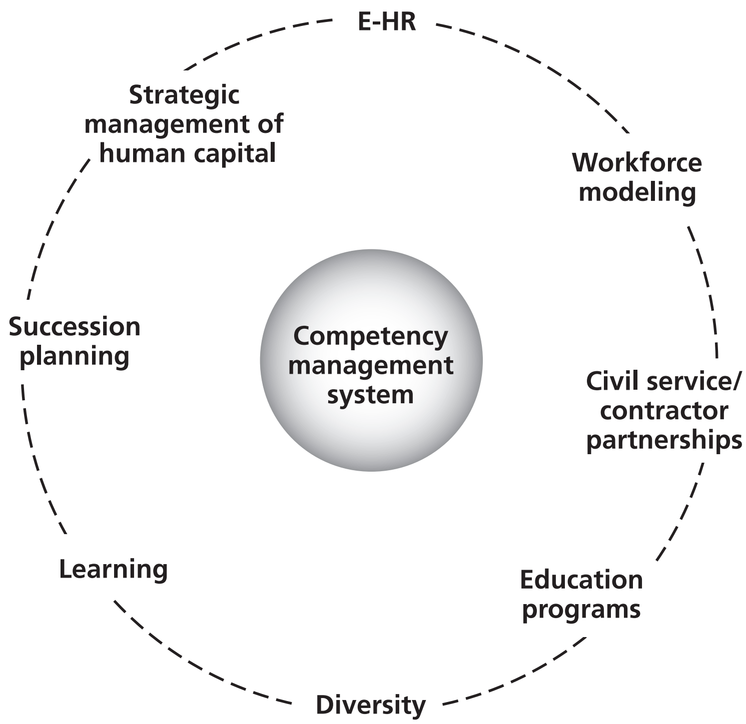
Use of Data in Human Capital Management Decisionmaking

Information such as that described in the last section can inform human capital strategy development and decisionmaking (see Figure 8.2).

For example, within NASA, the CMS enables the agency to compare future competency demands with the current knowledge base. When skills imbalances are identified, CMS information is tied into corporate recruitment, training, diversity, and education initiatives. For example, recent data analysis from CMS identified several competency areas that are "at risk" due to forecast attrition rates; this analysis generated the targeted areas for NASA's fall 2003 corporate recruitment activities. The CMS also provides critical linkages to mission goals, integrated financial, budget, and program management.

This ability to conduct detailed analysis of workforce requirements enables us to target future recruiting and outreach efforts as well. NASA has innovative programs to attract minorities and disabled persons and to create recruitment linkages between our employment programs and university research and education programs.

Figure 8.2
Integrated Workforce Information Processes



RAND *CF194-8.2*

Information from the CMS and external sources is also critical from the standpoints of retention and development of the workforce, and strengthening the education pipeline. By utilizing the workforce data to identify areas in which competencies must be strengthened, sound decisions can be made as to the best methods for addressing those needs—using existing tools or pursuing new tools through legislation, training and development programs, and leveraging external expertise from the private sector and academia. Workforce demographic data (age, attrition, retirement, etc.) also provide important information about the continuity of the agency's knowledge base and thus where resources should be focused to enhance knowledge sharing and mentoring of younger, less experienced workers.

With the challenges NASA faces, it is essential that the agency conduct effective workforce analysis and planning to align human resources with the agency's mission, goals, and objectives. The agency places great importance on the ability to obtain and analyze precise, comprehensive, and quantifiable data from which to develop focused STEM workforce strategies to ensure mission success today and in the future.

Types, Sources, and Improvements in Workforce Planning and Analysis Data—Some Opportunities

Understanding the above data sets can lead to a greatly improved and proactive decision-making process in other areas. A clear and consistent linkage of grants, fellowships, scholarships, and research opportunities across all federal STEM agencies could greatly assist the efforts of all organizations to develop, attract, and retain the "best and brightest" in the STEM areas of expertise. With limited federal resources among some agencies and limited private-sector resources for workforce development, a clear and coordinated road map for technology research, development, and application across the full range of STEM disciplines could enable an improved leveraging of the limited resources. A "road map" for technology development, designed, developed, and facilitated by the Office of Science and Technology Policy (OSTP), which would lay out current expectations and a future vision for technology development and application research, would enable colleges, universities, private industry, and federal agencies to consistently focus workforce development plans around overarching goals and the gathering of associated data for continual progress. Planning for the dissemination of various funding sources across multiple agencies could greatly improve the use of these funds for the taxpayer and enable agencies of various sizes the opportunity to help develop, or at least influence, the road map, and therefore to reap the benefits of the application of STEM disciplines.

It is perhaps essential that OSTP propose a consistent taxonomy and associated terminology that can be developed and used for appropriate comparison of STEM disciplines throughout their use in the workforce planning and analysis life cycle. For instance, knowing that a certain discipline is accurately and consistently understood across all colleges and universities enables a much better understanding of the health of the "pipeline" for talent and yields further insight into how those individuals can contribute to the success of various agencies with varied missions.

Finally, "share and share alike," as the saying goes. OSTP should enable and promote the consistent sharing of information in and among departments and agencies with particular interests in the STEM disciplines. This will enable the learning of new ways of recruiting and retaining new employees through finding what incentives and flexibilities work and which do not. It will enable learning in terms of development. For instance, programs to enhance the sharing of lessons learned and case studies can be more readily shared across agencies if the terminology at the STEM discipline level is somewhat flexible but consistent. Finally, the use of the various degrees, with consistent understanding of the trends associated with them, in a data format and structure understandable and consistent, can be shared in such as way as to bring to fruition the fullest use of the public trust that the federal government is applying resources to their fullest. The ability to share data associated with current as well as forecasted trends in STEM-discipline-related areas will be enhanced by consistency and will build the highly qualified and desired STEM workforce of the future.

Table 8.1
Competencies Representing NASA's Critical Needs: 4th Quarter, FY 2004 Analysis

	Category of Competency				
Competency	Corporate Level	At Risk	Shuttle Return to Flight	Fresh Out of College	NASA Center at Risk
Systems engineering	X	X	X	X	
Integration engineering	X	X	X	X	
Test engineering	X	X	X	X	
Mission assurance	X	X	X	X	
Human factors	X				
Nuclear engineering	X				
Design and development engineering	X	X	X		
Quality engineering and assurance	X		X		
Business management	X	X			
Mission execution	X	X		X	
Program or project management		X	X	X	
Business information technology systems		X			
Budgeting management		X		X	
Propulsion system and test			X		
Safety engineering and assurance			X		
Mission analysis and planning			X		
Administrative support (excluded)				X	
Acquisition and contract management				X	
Mathematical modeling and analysis				X	
Intelligent and adaptive systems					X
Institutional facilities planning					X
Institutional facilities operations					X
Fund human factors research					X
Engineering and science support					X
Electrical and electronics systems					X
Communications networks and engineering					X
Air-breathing propulsion					X
Computer systems and engineering					X
Financial management					X
Flight and ground data systems					X
Human resources					X
Public communications and outreach					X
Legal					X
Advanced mission analysis					X
Advanced experimentation and test technology					X
Aerodynamics					X

References

Green, Marnie, "Beware and Prepare: The Government Workforce of the Future," *Public Personnel Management*, Vol. 29, No. 4, 2000, pp. 435–444.

Grossman, Jerome H., "America Must Invest More Human Capital in Scientific, Technical Fields," op-ed press release, October 11, 2002. Online at http://www4.nationalacademies.org/onpi/oped.nsf/isbn/20021011?OpenDocument (as of June 4, 2004).

National Aeronautics and Space Administration, NASA's Strategic Human Capital Plan, 2002a. Online at http://nasapeople.nasa.gov/hcm/default.htm (current version online as of June 4, 2004).

_____, NASA Strategic Human Capital Implementation Plan, 2002b. Online at http://nasapeople.nasa.gov/hcm/default1.htm (current version online as of June 4, 2004).

_____, NASA Competency Management System, 2003a. Online at http://nasapeople.nasa.gov/workforce/data/page18.htm (current version online as of June 4, 2004).

_____, *NASA 2003 Strategic Plan*, 2003b. Online at http://www.nasa.gov/pdf/1968main_strategi.pdf (as of June 4, 2004).

_____, Workforce Planning, Web site, 2003c. Online at http://nasapeople.nasa.gov/workforce/data/ (current version online as of June 4, 2004).

National Science Foundation, *Science and Engineering Indicators 2002*, April 2002. National Science Board. Online at http://www.nsf.gov/sbe/srs/seind02/start.htm (as of June 4, 2004).

O'Keefe, Sean, "Seeking Scientists," GovExec.com, March 10, 2003. Online at http://www.govexec.com/dailyfed/0303/031003ff.htm (as of June 4, 2004).

U.S. Congressional Budget Office, *Changes in Federal Civilian Employment: An Update*, Washington D.C.: U.S. Government Printing Office, 2001.

CHAPTER NINE

Meeting the Data Needs: Opportunities and Challenges at the National Science Foundation

Lynda T. Carlson

Introduction

To propose recommendations for improving both data collection and the overall data system on the scientific and technical workforce for decisionmaking, one must understand what data exists. This paper provides a background on the data that the Division of Science Resources Statistics (SRS) at the National Science Foundation (NSF) collects on the scientific and technical workforce. The legal and policy origins of these data collections will be discussed. Of special interest will be the quality and statistical changes in the data sets that have occurred in the early 1990s and are occurring at present. These major changes were results of redesigns of the data collections to incorporate the results of the decennial census.

The paper addresses the wide range of activities under way in SRS to enhance our knowledge base on the scientific and technical workforce, as well as related areas. The paper highlights NSF's awareness and activities to improve its ability to provide estimates of foreign participation in the science and engineering (S&E) workforce, as well as the postdoctoral workforce. In conclusion, the paper notes areas of concern related to the S&E workforce that the Division of Science Resources Statistics has on its agenda to address.

Background to the Data Collections

Over 50 years ago, NSF, in its organic legislation, was given the mandate to

> [p]rovide a central clearinghouse for the collection, interpretation, and analysis of data on scientific and engineering resources and to provide a source of information for policy formulation by other agencies of the Federal Government. ...

Within NSF, that responsibility has been delegated to SRS, one of the 14 major statistical agencies within the federal government. To carry out this mandate, SRS designs, collects, and analyzes periodic data collections and surveys on the S&E workforce and research and development (R&D) performance, as well as a range of other activities, including compiling data from other sources, both nationally and internationally, on resources devoted to science, engineering, and technology. SRS disseminates the data it collects to policymakers and the public through a variety of vehicles. These include traditional reports of statistical tabulations, short analytical publications, longer special analytic reports, online data tabulation systems, and major analysis projects involving multiple dissemination mechanisms.

In addition, SRS has responsibility for producing the biennial Science and Engineering Indicators (SEI) publication for the National Science Board. This publication utilizes all the data series SRS collects and supports, as well as a broad range of additional primary and secondary data sources. In the very broadest sense, SEI can be considered the primary purpose of the ongoing SRS data collections of the S&E workforce, as well as R&D outlays and performance data collections. In the 30 years since the initial SEI, the range of data users and uses for the ongoing collections on S&E personnel has expanded markedly beyond that of the SEI. SEI remains the crucial driver for all the SRS data collections, and one of the major areas of interest in SEI is the characteristics, supply, and utilization of scientists and engineers in the components of the S&E workforce.

The second legislatively mandated requirement to NSF that is supported by NSF's S&E personnel surveys is a biennial report on the status of women and minorities in S&E. Unlike SEI, which covers the entire S&E research infrastructure, this report, entitled Women, Minorities and Persons with Disabilities in Science and Engineering, is a statistical compilation of information on the participation of women, minorities, and persons with disabilities in S&E. Beginning in 2004, the format will change from a report to an online database system with accompanying charts and graphics. These data will be updated systematically as new data become available.

The Survey of Earned Doctorates

A crown jewel of the data on scientists and engineers is the Survey of Earned Doctorates (SED), which NSF has been collecting for almost 50 years. Working with the degree-granting institutions, the NSF has collected basic demographic characteristics, information related to fields of study and field of degree, future work plans and study plans, and means of financial support received for the degree. This data set has served as a basis of much of our understanding of the doctoral workforce and the role of the United States in educating much of the world's doctorate recipients. The data set comprises doctorates awarded by U.S. doctoral-granting institutions. This includes both U.S. citizens and foreigners who are awarded doctorates by U.S. institutions.

Scientists and Engineers—The SESTAT System

Prior to 1993, NSF had in place a system called the Scientific and Technical Personnel Data System, which provided the core of the information on scientists and engineers. The system had as its core a database of individuals selected on occupation and as collected in the 1980 decennial census. These individuals were resurveyed every two or three years to update their information. In addition, separate surveys were in place to collect data on new graduates at the bachelor and master's levels who had received S&E degrees. Data on recipients of doctoral degrees were collected through the SED, from which a subsample of scientists and engineers was selected and resurveyed every two years in the Survey of Doctorate Recipients (SDR). For the most part, the surveys yielded low response rates and data that were often of questionable quality. These problems were the result of design issues, lack of adequate funding, and insufficient numbers of highly trained survey and mathematical statisticians working on the surveys.

In planning for a new sample from the 1990 decennial census, SRS commissioned the Committee on National Statistics (CNSTAT), National Academies of Science (NAS), to reassess the existing data system and specify the parameters for a new system (Citro and Kalton, 1989). The CNSTAT report was quite critical of the then-existing data system on scientists and engineers and quite explicit as to future design and needs if the nation was to have adequate information on this critical component of the workforce. One of the most critical components of the report dealt with how NSF then defined a scientist or engineer—through a highly complex computer algorithm, with points awarded to sample individuals based on occupation, education, and self-identification. This highly complex definition led to estimates of the number of scientists and engineers in the population that significantly diverged from other estimates then available from either the Census Bureau or the Bureau of Labor Statistics.

The CNSTAT report was critical of SRS methods of updating the data sets and the lack of any significant quality related activities. The panel was also critical of the inability of SRS to provide meaningful methods for potential users to use the data sets or to provide useable data sets for researchers. In reality, although NSF was supporting the collection of data, there was only limited data availability for researchers. This, combined with the then-limited analytical activities of SRS, undermined the very rationale for collecting the data or maintaining the databases.

The CNSTAT report strongly suggested that NSF adequately support significantly redesigned data collections on scientists and engineers and treat SRS as a federal statistical agency, with a primary mission to collect statistical data on scientists and engineering enterprise. The report was quite clear that inadequately designed and maintained S&E workforce surveys could not be continued and that a significant enhancement of resources was required. Starting in FY 1993, NSF did in fact undertake the changes indicated in that CNSTAT report and created the SESTAT system.

Utilizing the 1990 decennial census, SRS reengineered its workforce surveys along the lines proposed by the CNSTAT surveys. The new design produced integrated estimates of scientists and engineers based on a unified definition across all population groups. Beginning in 1993, the SESTAT data system, composed of three separate surveys, with an integrated definition of scientists and engineers, was instituted. SESTAT defines a scientist or engineer as either an individual who has received a college degree or higher in an S&E field or anyone who is employed in an S&E field. This definition has remained stable since 1993 and is the definition utilized for the 2003 data collections. What has been excluded, and will continue to be excluded, from this definition are technicians without college degrees. The unique part of the definition is that it is highly flexible. Researchers and analysts can drill down to different components of the data sets and create extracts for different types of analyses (Kannankutty and Wilkinson, 1999).

In implementing the recommendations of the CNSTAT report, SRS reengineered its surveys into three new workforce surveys with very distinct populations, then formed the SESTAT database. Each survey is conducted biennially through a decade in the odd years. The largest survey is the National Survey of College Graduates (NSCG)—this is the basic stock survey of scientists and engineers that is initially built off the decennial census. The initial NSCG survey of a decade is the one point in a decade that brings data on foreign trained or educated scientists and engineers who are in the U.S. workforce but not trained in the United States into the data set. The second is the National Survey of Recent College Graduates (NSRCG), conducted biennially, which provides a sample of the new stock of bachelor's and master's graduates to the system. SDR provides the third component. It is a longitudinal survey of U.S. S&E doctoral recipients. Over the past decade, SRS made continuous quality improvements to the surveys and data collection procedures and dramatically improved both the response rates and data quality.

The SESTAT surveys are significantly different from the surveys conducted prior to the CNSTAT report. They are completely integrated—from individual question wording, data collection procedures, and questionnaire printing to coding and editing specifications. Procedurally, the three surveys are completely compatible with each other. All three surveys use the same reference dates and are collected simultaneously. Naturally, each survey contains population-specific questions that are unique to the individual populations, such as doctoral recipients versus newly minted BS entrants to the S&E workforce. Data collection is in fact taking place right now and is quite a major operation to manage.

A second component of the improvements that SRS undertook in creating the SESTAT system was to integrate the data collected in SDR, NSCG, and NSRCG into a single database for the entire scientific workforce: the SESTAT Integrated Database. This is the database that SRS now supports through an online tabulation and extract system for analysts, policymakers, and the general public.

The SESTAT System for This Decade

In preparation for the 2000 decennial census, SRS undertook a detailed series of reviews, consultations, and evaluations of the existing survey design, the survey content, and present user needs. The intent of the redesign activity was to ensure that the data to be collected would be relevant to the policy and programmatic issues of the coming decade. An advisory group was established, and a series of workshops, interviews, and focus groups was held exploring the issues related to the data to be collected, the sample designs to be utilized, and even the visual design of the paper questionnaires and related materials.

The 2003 NSCG, which starts with a sample of approximately 177,000 college graduates from the decennial census, provides the core of respondents for the coming decade. This is the only point in the decade at which SRS is presently able to collect data on foreign trained scientists and engineers in the United States. Similarly, this is the only point in the decade at which we are able to obtain a baseline on individuals who are in S&E occupations but do not have S&E degrees. However, even at this point, it is not a complete estimate of scientists and engineers in the workforce, as foreign trained scientists and engineers or foreigners in S&E occupations who have enter the United States since April 2000 are not captured.

For the SESTAT surveys, the redesign evaluated both the core and module components of the three surveys. The core questions are asked for all three surveys and are collected in each round of the biennial of the data collections. The core questions enable SRS to produce the estimates of the S&E workforce size and demographic characteristics. The module components are special questions unique to each of the separate surveys, or subpopulations, and may be collected either in a single round of the survey or are asked only periodically.

Since the 2003 collections are the major decennial base year for the SESTAT surveys, they are critical to setting the framework for providing the trend data for the coming decade.

For 2003, the core questions are for the most part almost identical to those used in the 1993 survey. This is by design, as the NSCG will facilitate trend data across the decades. The core questions relate to demographics, education, occupation, income, and career patterns. The following changes were made to the core components of the SESTAT surveys:

1. A question on academic positions was added to all three surveys, and the tenure and rank questions were added to the NSCG and the NSRCG, to capture information on academic employment of non-PhDs and immigrants.
2. A complete degree inventory was added to the NSCG to obtain this information from new sample members from the 2000 decennial census.
3. Current enrollment or course taking was added to all the three surveys; this item was previously only asked on the NSRCG (which only surveys recent graduates), limiting the ability to fully understand continuing formal education.
4. For the NSRCG, the question on second or most-recent educational activity was revised to make the questions more clear.
5. The race and ethnicity questions were modified to meet the new Office of Management and Budget guidelines on collection of multiple-race data.
6. The SESTAT field-of-study list and the job code list were updated to adhere to the 2000 Classification of Instruction Programs, and the 2000 Standard Occupational Classification.

The following module items were either modified or fielded again for 2003:

1. The module on second jobs was removed (this item will only be fielded every other cycle).
2. Test questions that were first fielded in 2001 on job satisfaction and important characteristics of a job were added to all three surveys to gain a better understanding of these issues for all college graduates, since the 2003 NSCG includes a sample of individuals with a bachelor's or higher in non-S&E fields.
3. Publication and patenting activity questions were added to all three surveys. This item had been asked periodically in the past.
4. A special module was added to collect some limited information on immigrants. The types of questions include when they entered the United States, the reasons for immigrating, the type of visa they held, and dual-citizenship status.
5. For the SDR, the recent doctorates module was removed (this item will only be fielded every other cycle).
6. The SESTAT field of study list and the job code list were updated to adhere to the 2000 Classification of Instruction Programs, and the 2000 Standard Occupational Classification.

Improving the SESTAT Sample Designs

In preparation for the 2003 NSCG implementation, significant sample design work has been conducted by SRS staff and Census Bureau staff. Further, SRS commissioned a workshop by the Committee on National Statistics, NAS, to review potential design options considered by SRS (NRC, 2003).

The NSCG utilizes the decennial census long form information as the basic sampling frame. All individuals who have earned at least a bachelor's degree by the Census reference date are included in the basic frame. A sample of 177,000 individuals was selected. All degrees were represented, but an oversampling of S&E occupations and such S&E related occupations as secondary teachers and health fields was included. The ultimate target population is all individuals who have either an S&E degree or are employed in an S&E occupation (and to some extent the S&E related occupations). In addition, a unique component will be the retention of some of the longitudinal samples from the previous decade, especially those members of the S&E workforce in the sample who are underrepresented minorities. This longitudinal component is approximately 40,000 cases.

The new design will enable SRS to provide significantly more data on scientists and engineers by gender, race and ethnicity, major degree field, and immigration status. Furthermore, the retention of individuals from the 1993 NSCG will provide researchers with the ability to conduct longitudinal analyses across the decades, a unique feature that was not previously available, and to make quality assessments previously not available.

A significant redesign also took place with the NSRCG. Over time, the ability of this survey to provide detailed information on race, ethnicity, and sex of S&E graduates at the bachelor's and master's levels had declined. To improve the survey, the sample size was significantly increased, from 13,500 to 18,000. Coverage for the health fields has been newly added to the 2003 NSRCG and will be maintained in the future. Efforts have been made to increase the coverage of minority recipients of S&E degrees.

Informing the S&E Workforce Policy Issues

The function of a federal statistical entity is not to collect data for data's sake, but rather to be knowledgeable concerning the policy issues in areas of concern and then to collect the data that will inform the policy debate. For NSF, the strategic goals are people, ideas, and tools—to advance the frontiers of science, build a world-class scientific workforce, and develop the necessary tools and facilities to enhance the productivity of the research and education enterprise. That is, in fact, the function of SRS: to collect the information that will inform the policy debate with respect to the S&E enterprise. The SESTAT system, a component of the SRS suite of surveys, is designed to inform the policymakers and others with respect to issues concerning the S&E workforce.

NSF makes significant investments in people to develop world-class scientists and engineers, as well as a national workforce that is "strong scientifically, technologically and mathematically." Scientists and engineers are critical to a nation's economic productivity; hence, an understanding of what is occurring with respect to that workforce is critical. Specific issues of concern tend to relate to the following questions:

- What is the stock of scientist and engineers?
- What is the supply of S&E graduates?
- What are the training and the skill sets of scientists and engineers of both the existing stock and new scientists and engineers?
- How can we attract more U.S. citizens, especially women and minorities, to S&E?
- What is the role of foreign citizens in the U.S. S&E enterprise, and is that changing?

- What is the role of international mobility of scientists and engineers in the U.S. S&E enterprise?
- How and by whom is research being conducted in both academe and industry?
- What is the role of graduate students and postdoctorates in the research enterprise?
- What are the career prospects for scientists and engineers in academe, industry, and the nonprofit sectors?
- What is the expected impact of retirements and demographic shifts on the S&E workforce?
- What are the degree trends by demographic groups?
- What is the academic employment of scientists and engineers?

Do NSF Data Systems Inform the Policy Issues? What Are the Challenges for This Decade?

There are a number of "grand challenges" in the S&E workforce area that SRS faces over the next few years. As data on the S&E workforce become even more critically important, the existing data gaps and challenges must be addressed.

Examining both the policy issues and the SRS data collections, it is evident that the existing data collections can only shed light on components of these policy issues. For instance, the most complete estimate of the stock of scientists and engineers in the United States can only presently be captured at the start of the decade, when the NSCG is built from the decennial census and is able to capture an estimate of the immigrant scientists and engineers in the population. Yet even here, SESTAT is missing foreign-degreed scientists and engineers who entered the workforce after April 2000. It is only in the first NSCG of the decade that the stock estimate includes individuals without S&E degrees who are working in S&E occupations. Depending on the state of the economy, both of these categories many lead to either substantial under- or overestimates of the stock of the S&E workforce as the decade progresses. SRS will be examining the possibility of using immigration and visa samples and possibly administrative records to improve both the SESTAT surveys and data estimates throughout the decade. It will be critical for SRS to deal with this data gap. This issue will be a component of the SRS workplan for the next several years and is likely to require significant resources.

The SESTAT system, with the biennial input of doctoral, master's, and bachelor's graduates, is able to produce high-quality estimates of U.S. S&E graduates. Unfortunately, since similar surveys do not exist internationally, except in a few select countries, international estimates of S&E graduates do not exist.

The SESTAT system, with its excellent demographic data set, is capable of providing excellent estimates of graduates by a broad range of demographic characteristics, including gender and race. However, there are limitations on the data quality when information is requested on individuals with disabilities. Similar problems occur when information is requested on national origin of foreign students—for many fields or countries, the sample sizes in SESTAT are too small to provide information by highly detailed categories. These limitations became quite apparent after September 11, 2001, when SRS was often unable to respond to requests for summary statistical data that required, for example, nationality of foreign graduates crossed by detailed field of study.

Data on the training and skills sets of scientists and engineers within the SESTAT population are known. As respondents to the NSCG and SDR are resurveyed biennially, information on new training and education is collected. However, data are missing on foreign-degreed scientists and engineers who enter the U.S. workforce after the fielding of the decennial census or individuals who function in S&E occupations without S&E degrees. In addition, the present ability to capture interdisciplinary educational patterns of graduates at all levels has been somewhat limited but will be substantially improved in the coming decade. Related to information on interdisciplinary experiences is data on training in "teaming" and the ability to work in groups. Increasingly, that is a policy question that existing data are unable to answer.

Only to a very limited extent can the SESTAT system provide any information related to attracting more U.S. citizens to S&E. The demographic and educational patterns of new graduates in S&E provide no information related to cohorts who either dropped out of S&E educational streams or left S&E positions.

Information on the role of foreign citizens in the U.S. S&E enterprise is probably the foremost policy question related to today's workforce. SESTAT can provide excellent information at the beginning of the decade and on any foreign citizen who receives a degree in the United States. For those who receive a doctorate and remain in the United States, we can and do follow their career patterns over the decades. However, what is critically missing is detailed statistical, rather than anecdotal, information on the role of foreign doctorates in the U.S. research establishment. Critically important components are the foreign-degreed scientists and engineers who come to the United States to continue their research or education as postdoctorates in U.S. research universities. From the SRS Survey of Graduate Students and Postdoctorates in Science and Engineering, estimates are available on the number of foreign postdoctorates by academic institution and department. SRS is aware that these numbers are undercounts of the actual number of postdoctorates in academe and that no data are collected on postdoctorates in nonprofit institutions or industry. In fact, SESTAT is limited with respect to data on postdoctorates, who are a critical component of the S&E workforce.

Surveys of individuals who are either graduate students or postdoctorates in S&E fields are not conducted by SRS. Only limited aggregate data on postdoctorates and graduate students are collected. Specific information relates to fields of study, demographics, and means of support.

International mobility of U.S.-educated scientists and engineers is becoming critically important as countries are establishing programs to actively recruit and attract back their citizens who have received degrees in the United States. Equally important is to understand the international mobility, both for education and careers, of U.S. citizens. SESTAT has in the past included modules on international experiences of respondents. However, neither the NSRCG nor the SDR located or maintained in sample any respondents who indicated that they were going overseas either for employment or educational opportunities. Starting with the 2003 NSRCG and SDR surveys, experiments are under way to determine whether we can both track and collect data from these individuals.

Some of the "grand challenges" with respect to the S&E workforce relate to industry and how the S&E workforce is deployed in that sector. The Industry R&D survey conducted by the Census Bureau for SRS is substantially limited in its ability to provide data on the role of scientists and engineers in industry and how research is conducted in that sector.

The Industry R&D survey provides only limited aggregate information on researchers in industry.

The SESTAT surveys provide data on patents and publications by scientists and engineers but provide no information on how research is actually conducted. The existing taxonomies used in the SESTAT and R&D surveys are in need of substantial revision to keep up with the continual development of new fields. The existing SESTAT surveys provide no data on the role of graduate students and postdoctorates in various research enterprises beyond aggregate numbers of each category and their forms of support. The growth of interdisciplinary and multidisciplinary research and teaming across disciplines, departments, or centers is not captured in the existing surveys. The role of teams, be it within a single institution, across institutions, or even internationally, in research endeavors is difficult to capture within the constructs of the existing surveys. Yet NSF, as well as other funding institutions, is encouraging, through cross-disciplinary initiatives and various center solicitations, the development of nontraditional methods of conducting research.

Unlike many of the Organisation for Economic Co-operation and Development (OECD) countries, the SRS survey of Industrial R&D collects only limited data on the role of scientists and engineers as part of its annual collection.

SRS as a statistical agency collects data and analyzes the data that it collects. SRS also utilizes a broad range of other information on the S&E enterprise. However, SRS does not forecast or model future career prospects, projected impacts of retirements, or demographic shifts. The data that SRS collects facilitate the ability of researchers and other government agencies with responsibility for forecasting to make these projections.

SESTAT is well suited to provide data on degree and field trends by demographic groups. However, similar concerns exist with respect to the ability of SESTAT and SED to accurately assess the development of emerging fields and interdisciplinary fields and the production of graduates in these fields.

SESTAT can provide part of the picture of academic employment of scientists and engineers. Because the SDR is a microdata file, very detailed information related to types of appointments, teaching and research responsibilities, salaries, benefits, and related information are obtained. However, these data are only available on individuals who received their doctoral degrees in the United States. Hence, data on foreign-degreed academicians, a substantial component in many fields, are sorely lacking.

Improvements Under Way to Enhance the Capabilities of NSF Data Collections

As a statistical agency, SRS is very well aware of both the deficiencies in the existing surveys and the challenges it faces in building new surveys, improving existing surveys, and maintaining a continuous quality-improvement process. NSF has a strong commitment to the improvement and enhancement of the SRS statistical data collections, which serve as the base of knowledge on the S&E enterprise. NSF has made a multimillion-dollar commitment to improving the SESTAT data collections for this decade. It is expected that these efforts will lead to major changes in the SRS portfolio of surveys. A number of activities are already under way in SRS to improve both existing and planned data collections related to the S&E workforce.

To have a better understanding of international mobility of the highly educated, SRS has instituted new procedures. For the 2003 fielding of the NSRCG, significant attempts will be made to locate sample respondents (many of whom may be foreign students) who have left the United States. Once located, the complete survey will be conducted with these respondents. A special question has been added for these respondents to determine whether they have plans to return to the United States. For the longitudinal SDR survey, respondents leaving the United States were not, in the past, maintained in the sample. A methodological test is now under way, starting with the 2003 SDR, to see whether it is feasible to recapture prior-year respondents who had left the country, as well as potential respondents from the 2002 SED who indicated that their next positions would be outside the United States. If these experiments are successful, inclusion of respondents who have left the country will be included in both the NSRCG and SDR surveys in the future. These improvements will greatly enhance the ability of both NSF and researchers to understand the international mobility patterns of highly trained scientists and engineers who have been trained in the United States and then left the country, either permanently or temporarily.

A greater understanding of the international mobility of scientists and engineers suffers from a lack of international data on the supply of doctorate holders, as well as their deployment. There is a broad range of policy questions, both in the United States and internationally, which are ill informed because data internationally on doctorate holders do not exist. The range of questions relates to the supply of doctorate holders, the supply of foreign doctorate holders, industry employment, and involvement in doctoral and beyond research, among others. This information becomes even more interesting and important as research projects are conducted through international collaborations—be it within a single firm, between firms and academe, and/or between teams of academic researchers in various institutions.

To better understand the supply, role, and career patterns of doctorate holders, SRS has begun working with Statistics Canada, OECD, and the United Nations Educational, Scientific, and Cultural Organization's Institute of Statistics (UIS) to foster the development of doctoral data sets in other countries—particularly countries that have been traditional suppliers to the United States. SRS has provided substantial technical expertise to Statistics Canada in the development of the Canadian SED, which was successfully piloted this year. A full-scale, ongoing Canadian SED is planned for next year and will be conducted annually thereafter. SRS has been working with the OECD to encourage the development of internationally comparable data on doctorate holders in OECD countries. Because only a very few countries collect data on doctorate holders, OECD will be holding a series of workshops on the potential of collecting these data and then undertaking pilot surveys in interested member countries. A similar project is planned for UIS, with SRS support, and again pilot surveys will be conducted in countries that are not part of the OECD system. The two projects will be coordinated, and SRS will be directly involved and is providing initial seed funds. It is hoped that these activities will lead to ongoing doctorate surveys from a significant number of countries in the future.

Compared with those of other OECD countries, the U.S. surveys of research and development funding in industry are severely limited with respect to identifying numbers and types of scientists and engineers linked to research expenditures by individual establishments. SRS is beginning a major review of the Industry R&D survey collected by the Census Bureau for NSF. A response analysis survey, to explore the ability to collect this data from industry and potential new questions, is planned for next year. One area of exploration will be the S&E workforce for individual companies and whether accurate data on counts and professional background are obtainable.

SRS is attempting to improve its information on secondary teachers of science. For the 2003 NSCG, we oversampled S&E and S&E-related occupations. If the NSCG response rate is sufficiently high, this should permit a detailed analysis of the degree background and career patterns of secondary science teachers in the S&E workforce.

SRS also worked with the National Center for Education Statistics to add a question on field of study to the October 2002 education supplement of the Current Population Survey (CPS). The CPS monthly surveys are cross sections of the general population. SRS will be analyzing these data to determine whether the CPS is a feasible means of obtaining information on S&E workers who have less than a baccalaureate degree. The data collection also serves as a first methodological test of whether SRS can utilize the American Communities Survey as the sampling frame for the NSCG. This will be an important analysis, because the long-form questionnaire for the Census, which serves as the universe for the NSCG, is not expected to be collected in 2010 but rather be replaced by the American Communities Survey.

To better capture interdisciplinary and multidisciplinary degrees and activities and S&E degree holders, SRS has begun a series of activities aimed at improving existing survey questions on fields of study and fields of degree. A series of methodology tests was carried out on how best to restructure these questions on the SED and the SESTAT surveys to get a better estimate on the growth of interdisciplinary degrees and fields. SRS has also begun a multiyear activity to improve its taxonomies. It is expected that there will be substantial enhancements to the taxonomies used for both the workforce and the R&D surveys. Part of the activity is to improve the concordance between the taxonomies used for the R&D surveys of performance and federal funds with the SESTAT and related surveys.

As indicated earlier, postdoctorates are a major component of the training of scientists and engineers, as well as a significant part of the S&E research workforce. A postdoctorate is a position generally held by a recently awarded PhD that involves full-time research. The actual number of known postdoctorates in all science, engineering, and health fields in 2000 was 41,548. More than half these numbers are temporary visa holders. It has long been known that the data NSF collects on postdoctorates result in an undercount, missing postdoctorates in industry, the federal government, and nonprofits. Further, the existing survey appears to miss a significant number of individuals who are paid through grants received by principal investigators.

The existing SRS survey on graduate students and postdoctorates is based on institutions rather than individuals and collects only aggregate counts of postdoctorates by department, with information on sources of support and citizenship, which is extraordinarily limited information on this significant component of the scientific workforce. Starting with this fiscal year, SRS has initiated a comprehensive study of the feasibility of developing a new survey of individuals to collect valid, useful, and reliable information about individuals in postdoctoral positions. Planned activities include assessing information needs about this population, including emerging concerns related to international migration and national security, as well as heightened concerns about compensation, working conditions, and professional development. Also to be examined will be the types of positions and appointments to be included in a survey, as well as possible approaches to developing a sampling frame for these positions across all employment sectors—academia, government, private for-profit, and nonprofit.

Foreign Scientists and Engineers in the United States: The Challenge of Keeping the Data Current

Once a decade, with the decennial census sample, the SESTAT system has almost complete information on foreign scientists and engineers in the United States. What is missing from the count are any scientists and engineers who entered the United States after the decennial census for employment or a postdoctoral opportunity and non-S&E degree holders working in S&E occupations. However, as a decade progresses, the ability of the SESTAT system to accurately estimate the number of foreigners in the workforce declines because we do not presently have a means of updating the inflow of S&E workers or postdoctorates. As part of the development of the postdoctorate survey, this undercount issue will be directly addressed. One potential means of updating both the SESTAT workforce estimates and sampling postdoctorates could be the Student and Exchange Visitor Information System (SEVIS). Over the next several years, SRS will be exploring the potential of SEVIS, as well as other schema to provide lists of foreigners coming to the United States for postdoctoral opportunities. SRS will be exploring both the ability of SEVIS as a quality check on the existing postdoctoral data and the sampling frame potential. SEVIS may prove to be the most efficient means of obtaining high-quality current data on foreign students and postdoctorates on an ongoing basis.

The Challenges

Over the past ten years, SRS has made significant improvements in its data collection and analysis on the S&E workforce. The challenge to NSF for the future will be quite broad and will include taxonomies, postdoctorates, and the S&E workforce in industry. By far the greatest challenge will be to develop a means of keeping current the data on foreign citizens in the S&E workforce. SRS has proven that it can make significant improvements to its surveys if sufficient resources are available.

References

Section 4(j)1, National Science Foundation Act of 1950, as amended.

Public Law 95-516, Science and Engineering Equal Opportunities Act, 1980.

Citro, Constance F., and Graham Kalton, eds., *Surveying the Nation's Scientists and Engineers: A Data System for the 1990s*, Washington, D.C.: National Research Council, 1989.

Kannankutty, Nirmala, and R. Keith Wilkinson, *SESTAT: A Tool for Studying Scientists & Engineers in the United States*, Arlington, Va.: National Science Foundation, NSF 99-337, 1999.

National Research Council, *Improving the Design of the Scientists and Engineers Statistical Data System (SESTAT)*, Washington, D.C.: The National Academies Press, 2003.

National Science Foundation, FY 2004 Budget Request to Congress.

National Science Foundation, Division of Science Resources Statistics, *Graduate Students and Postdoctorates in Science and Engineering: Fall 2000*, Washington, D.C., NSF 02-314, 2000.

National Science Foundation, Division of Science Resources Statistics, *Women, Minorities and Persons with Disabilities in Science and Engineering: 2002*, September 2003. Online at http://www.nsf.gov/sbe/srs/nsf03312/start.htm (as of June 4, 2004).

CHAPTER TEN

Opportunities and Challenges at the Bureau of Labor Statistics

Michael W. Horrigan[1]

As phrased, the subject of the conference—Hw will we recognize a shortage or surplus in the scientific and technical workforce: Improving the data system for decisionmaking?—suggests that determining whether or not a labor shortage (or surplus) exists can be highly problematic. This paper agrees with this implicit assumption and argues that the difficulties result from both the traditional empirical approach taken to measuring shortages (surpluses) and to the substantial gaps in knowledge about unfilled vacancies by occupation. An alternative, admittedly more cautious, approach is developed and is applied in a case study format to six different types of engineering occupations: mechanical engineers, aerospace engineers, industrial engineers, electrical and electronic engineers, civil engineers, and chemical engineers. In what follows, labor shortages will be used as the anchor for the analysis, leaving the discussion of labor surpluses as the implied analogous case.

The classic theoretical definition of a labor shortage, the persistence of vacancies even in the face of rising wages, implicitly describes a labor market in which the demand for labor is rising faster than the supply that is forthcoming in response to higher wage offers. The traditional approach to determining whether a labor shortage exists is to model the demand for and the supply of labor separately and use the difference between the two as a measure of current or, as relevant, future labor shortfalls. Such a seemingly straightforward task immediately poses a number of challenges. First, what is the unit of observation? Are vacancies being defined on the basis of detailed occupations, or are vacancies being defined on some other dimension, such as worker skills? Second, what is the unit of time reference? Is the policy concern over current or future labor market conditions? If the latter, then is the perspective one of a relatively short-term nature, such as the next three years, or is it long term, say ten years or more?

If one could confidently measure the current or future demand for and the supply of workers by detailed occupation—much less by skill level—then one could also inform policymakers as to the potential for worker shortages by these same dimensions. However, the difficulty of measuring current or projected demand and supply by occupation or by worker skills presents a formidable obstacle. Currently, no federal statistical survey provides measures of vacancies and employer wage offers by detailed occupational group.[2] Measures of the supply and demand for worker skills are even more problematic.

---

[1] Michael W. Horrigan is Assistant Commissioner in the Office of Occupational Statistics and Employment Projections at the Bureau of Labor Statistics, Washington, D.C.

[2] The Bureau of Labor Statistics (BLS) Job Openings and Labor Turnover Survey provides estimates of job openings by industry for the entire U.S. economy. Twelve states have conducted job vacancy surveys that provide estimates of job openings by occupation.

Another obstacle is the difficulty of modeling, for each detailed occupation, the dynamic labor market responses to shortage conditions. In addition to changing wage offers, there are numerous other possible adjustments that employers may adopt when they are having a difficult time finding workers. How industries manage their human resource requirements is influenced by a great many factors, including, but not limited to, the available labor supply, including immigration; the skill levels of prospective job seekers; the use of technology in the production process and the required capital-labor ratio consistent with the technology used for production; how work is organized; the use of employees from the personnel supply services industry; the hiring of self-employed contractors; the use of flextime and "flexiplace"; the use of overtime or mandatory shift coverage; and the hiring of offshore labor in other countries.

This paper argues for an equilibrium approach to determining whether current labor market conditions for detailed occupations are consistent with evidence of a labor market shortage. In particular, is there a pattern of substantial increases in nominal wage offers that elicit or are accompanied by fairly small increases in net employment levels—indicating difficulties in finding sufficient supplies of skilled workers in the occupation?[3] Certainly, this is not the only circumstance in which an occupation may be in shortage. For example, a shortage can exist in the short run but be effectively eliminated if rising wages result in substantial increases in employment over longer periods. Alternatively, an occupation can be thought of as being in a shortage condition even when wages are not rising in an attempt to restore labor market equilibrium. In particular, it is possible that employers are adjusting at other margins, such as hiring immigrants, offering or requiring overtime, hiring temporary workers, or moving jobs offshore, to name a few.

While such pure nonwage adjustments may be observed in response to hiring difficulties, the approach taken in this paper to identifying shortages is stricter and admittedly more traditional. Substantial increases in nominal wages that are accompanied by weak employment growth are used to initially screen occupations for the possible existence of long-term shortage conditions. A second screen that is applied asks whether the fortunes of a given occupation are closely tied to the fortunes of particular industries or a set of industries. And if so, how have these particular industries fared over the period under analysis? Does the profile of output in these industries suggest there has been continued strong demand for the services of employees in the occupation?

This triage approach narrows the list of occupations being considered as having long-term shortage conditions. For the remaining occupations, other complementary pieces of evidence are examined to determine whether, if combined, they paint of portrait of a labor market for detailed occupations that are facing long-term shortages. This examination includes asking whether the experienced unemployment rate for these occupations was low (or whether it got significantly lower), indicating a relatively small untapped labor force willing and ready to take a job. Did employers seek to expand their base of employees, reaching out to various demographic, immigrant, or armed service veteran groups that may not be typical employees in the occupations? Did employers offer more overtime or increase the use of part-time workers? Did educational attainment requirements for the occupation

[3] This is an equilibrium approach in the sense that it is only the outcomes of the interactions of the supply-and-demand process that are measured.

decline over the period, as evidenced by a significant leftward shift in the educational distribution of the occupation?

While this approach provides evidence on possible shortage occupations based on past data, what can be said about prospects for future labor market shortages? It is recommended that analysts look at a number of indicators. What is known about projections of net employment growth in occupations? For example, the BLS program of long-term employment projections provides estimates of net employment change for 10 years in the future. Is the occupation projected to grow faster than average? And are there significant anticipated replacement needs for the occupation?

As mentioned above, the market for six different engineering occupations will be examined. The purpose of this analysis is to demonstrate, in a case-study format, the use of existing data to build a prima facie case as to whether there are substantial current or anticipated difficulties in hiring workers from selected occupations. Data from the Current Population Survey are used to examine the pattern of nominal wages and employment. In particular, the 1994–2000 period is used to compare labor markets at the early stage of a job market recovery (1994) to those at the end of an expansion (2000). As necessary, to guarantee sufficient weighted cell sizes for estimation, data for 1994–1996 are combined and compared to data for 1998–2000. Otherwise, comparisons are made between 1994 and 2000. Finally, in addition to examining the historical employment and wage trends, BLS projections of occupational employment for the 2002–2012 period are used to examine the anticipated demand for these occupations in the long run.

Table 10.1 provides the evidence on employment and wage trends over the 1994–2000 period. Three occupations—mechanical engineers, aerospace engineers, and industrial engineers—experienced substantial increases in nominal wages over the period, but the change in labor supply in each of these occupations associated with these higher wages, while positive, was extremely weak. Wages for mechanical engineers and aerospace engineers rose faster than the average increase for all occupations, while wage increases for industrial engineers were only slightly less than the overall average. The labor market in all three of these occupations had a very sluggish labor supply response associated with the significant increases in wages, indicating labor markets in which there may be long-term difficulties finding workers. In contrast, three occupations—electrical and electronic engineers, civil engineers, and chemical engineers—each experienced substantial increases in employment in association with wage increases. In particular, for electrical and electronic engineers and for civil engineers, if shortage conditions existed, they were temporary and were effectively erased by the pattern of significant increases in labor supply in response to the rising wages. The sixth engineering occupation, chemical engineers, experienced a highly elastic labor supply response in association with higher wages. In this case, a 4.6 percent nominal wage change between 1994 and 2000 was accompanied by a 65.3 percent net supply increase of full-time workers in this occupation between 1994 and 2000.

Of the six occupations discussed, only three pass the initial screen for possible long-term difficulties in finding workers: mechanical engineers, aerospace engineers, and industrial engineers. The next screening test assesses whether the fortunes of any of these occupations are tied to the fortunes of a particular industry or set of industries—and how have these industries fared? One of these occupations, aerospace engineers, is concentrated in two aero-

Table 10.1
Employment and Nominal Wage Change for All Workers and for Selected Engineering Occupations, 1994 to 2000

	Employment (000)			Average Nominal Earnings ($)			Average Real Earnings (in 2000 CPI-U adjusted $)		
Occupation	**1994**	**2000**	**% Change**	**1994**	**2000**	**% Change**	**1994**	**2000**	**% Change**
Wage and salary workers who usually work full time									
Total, all workers	87,379	99,917	14.3	574	715	24.6	697	724	3.9
Mechanical engineers	321	325	1.4	937	1,183	26.2	1,139	1,198	5.2
Aerospace engineers	73	77	5.7	1040	1,305	25.5	1,263	1,322	4.6
Industrial engineers	239	245	2.5	859	1,044	21.5	1,043	1,057	1.3
Electrical and electronic engineers	525	687	30.8	1,013	1,240	22.5	1,231	1,256	2.1
Civil engineers	203	258	27.2	977	1,122	14.9	1,187	1,137	–4.3
Chemical engineers	53	88	65.3	1,307	1,368	4.6	1,589	1,387	–12.8
All employees									
Total, all workers	107,988	120,786	11.9	500	632	26.4	608	640	5.3
Mechanical engineers	324	334	3.1	933	1,166	25.0	1,134	1,181	4.2
Aerospace engineers	75	77	3.9	1042	1305	25.3	1,266	1,322	4.4
Industrial engineers	246	248	0.9	842	1037	23.1	1,023	1,050	2.6
Electrical and electronic engineers	535	705	31.7	1,001	1,222	22.1	1,217	1,238	1.8
Civil engineers	209	267	27.7	957	1,099	14.8	1,163	1,113	–4.3
Chemical engineers	53	90	68.7	1,307	1,361	4.1	1,589	1,379	–13.2

space industries. In 2000, 61 percent of aerospace engineering employment was in the aircraft and parts and the guided missiles, space vehicles, and parts industries. Real output in these two industries declined over the 1990–2000 period, from $142 billion to 136 billion (measured in 1996 chain-weighted dollars). The other two occupations are spread out over more than 100 industries. In the case of mechanical engineers, no one industry employed more than 8 percent of employees in this occupation group. One industry, engineering-architectural-surveying, employed 15 percent of industrial engineers, while no other industry employed more than 6.5 percent.

The concentration of aerospace engineers in a declining industry makes it difficult to establish that the small employment gain in this occupation from 1994 to 2000 was the result of an occupation in a long-term shortage condition. Indeed, the declining fortunes of the aerospace industry may have discouraged entry into the aerospace engineering occupation over this period and may quite possibly have made it difficult to find workers in this occupation—leading to the observed increases in nominal wages. In addition, the 2002–2012 BLS employment projections estimate a net employment decline of 5 percent for this occupation. Taken together, these pieces of evidence suggest ruling out this occupation from further consideration.

For the remaining two occupations—mechanical and industrial engineers—a natural next step is to examine whether there were other labor market responses that would reinforce the view that there have been long-term difficulties in finding adequate supplies of workers. Table 10.2 provides evidence on the experienced unemployment rate[4] for these two occupations. The experienced unemployment rates declined steadily for both engineering occupations and were consistently below the rate for all workers. A relatively low experienced unemployment rate is consistent with the idea that workers in these fields might become difficult to find if the demand for their services were to expand significantly in the short run.

Table 10.3 provides the evidence on a variety of other characteristics. As the table indicates, both engineering occupations have extremely small shares of female employees, a

Table 10.2
Experienced Unemployment Rate for All Workers and for Mechanical and Industrial Engineers, 1994 to 2000

Year	All Workers	Mechanical Engineers	Industrial Engineers
1994	5.7	3.7	2.6
1995	5.2	1.8	2.3
1996	5.0	1.5	2.5
1997	4.5	1.3	1.6
1998	4.1	1.3	2.1
1999	3.9	2.0	2.0
2000	3.7	1.3	1.3

[4] The experienced unemployment rate for a specific occupation is based on the concept of experienced unemployment in the occupation—that is, the number of unemployed individuals whose last job was in that occupation. The experienced unemployment rate is the number of experienced unemployed in the occupation divided by the sum of employed and experienced unemployed in the occupation.

Table 10.3
Percent Distribution of All Workers and Mechanical and Industrial Engineers, by Selected Characteristics, 1994–1996 to 1998–2000

Characteristics	All Employees 1994	All Employees 2000	Mechanical Engineers 1994	Mechanical Engineers 2000	Mechanical Engineers 1994	Mechanical Engineers 2000
Gender						
Male	53.9	53.6	94.5	93.2	85.8	83.2
Female	46.1	46.4	5.5	6.8	14.2	16.8
Age						
16–24	15.1	15.1	3.5	4.3	4.2	4.5
25–54	72.8	72.1	81.5	83.0	82.9	83.3
55 and older	12.2	12.9	15.0	12.8	12.9	12.3
Race/ethnicity						
Hispanic	9.0	10.4	3.6	2.7	4.5	4.3
Non-Hispanic black	10.4	10.9	3.5	3.2	4.8	5.3
Non-black non-Hispanic	80.6	78.7	92.9	94.1	90.6	90.4
Hours worked						
Less than 40 hours	31.3	29.0	6.7	6.7	6.4	6.9
40–44 hours	48.5	51.1	58.9	57.5	55.0	55.7
45–49 hours	5.9	5.7	12.2	14.7	14.3	13.9
50 hours or more	14.3	14.2	22.3	21.2	24.3	23.6
Part-time status						
Works part time	19.8	18.5	2.2	2.6	2.1	2.0
Does not work part time	80.2	81.5	97.8	97.4	97.9	98.0
Veteran status						
Served in the armed forces	12.5	10.5	19.4	16.1	26.0	19.0
Did not serve in the armed forces	87.5	89.5	80.6	83.9	74.0	81.0
Immigration status						
Did not immigrate to the United States	88.7	86.7	85.7	85.8	91.0	90.0
Immigrated within last 5 years	1.6	1.5	1.6	0.9	1.3	1.1
Immigrated 6–10 years ago	2.2	2.6	2.1	3.0	1.1	2.0
Immigrated more than 10 years ago	7.5	9.2	10.7	10.3	6.6	7.0
Educational attainment						
High school graduate or less	45.6	44.3	6.5	6.3	13.5	12.0
Some college	28.7	28.5	21.1	19.7	26.6	24.7
College degree or higher	17.2	18.3	72.4	74.0	59.9	63.3

characteristic that did not change over the 1994–2000 period. Nor did the race and ethnicity shares change substantially across the period. The proportion of mechanical engineers working 45 to 49 hours per week rose from 12.2 to 14.7 between 1994–1996 and 1998–2000. However, for both mechanical and industrial engineers, the overall proportions of individuals who work 40 hours a week or more was relatively stable over the period. Nor was there any significant shift in the proportion of workers who work full or part time, the latter being a very small percentage of the total for both occupations over the period. Reliance on employees who have served in the military actually decreased in relative terms, and the proportions of employees who are immigrants was virtually unchanged for both occupations over the period. And finally, the evidence on educational attainment shows a trend toward the preferred hiring of 4-year college graduates for both of these occupations, since the percentage of those with some college fell and the percentage with 4 years of college or more rose over the period. The lowering of relative skill requirements needed to gain entry into each occupation does not appear to have been a deliberate strategy on the part of employers. Overall, for both occupations, the evidence does not support the idea that firms, in responding to long-term shortage conditions, adopted any significant nonwage adjustments in their hiring or workplace practices to increase either the number of employees or the number of hours worked.

What about future prospects for these three occupations? What is the expected increase in net employment for these occupations? Table 10.4 provides the BLS projections of occupational employment for the 2002–2012 period. Net employment in both of these occupations is projected to increase over the period, although the percentage change is less than the percentage change for all occupations.

Another potential indicator of future shortages is the number of workers in the occupation who will need to be replaced owing to retirements. Table 10.3 provided the shares of employment in each occupation aged 55 and older. Overall, the percentage of employees aged 55 and older rose from 12.2 to 12.9 percent between 1994–1996 and 1998–2000. In contrast, the percentage of mechanical engineers aged 55 and older fell from 15.0 to 12.8 percent, and the percentage for industrial engineers also fell, but only slightly, from 12.9 to 12.3 percent.

What does the overall evidence suggest in terms of long-term difficulties in finding sufficient supplies of mechanical and industrial engineers? There does appear to have been a deliberate rising wage strategy that was relatively unsuccessful in attracting new workers into these occupations over the 1994–2000 period. In addition, BLS projections for 2002–2012 indicate continued increases in demand for the services of these workers. In contrast, however, the evidence does not indicate that employers were adopting various alternative nonwage strategies for hiring workers in these two occupations. And perhaps the single most important piece of evidence missing to bolster the view that these occupations suffered long-term difficulties in finding skilled workers over the 1994–2000 period is information on the number of vacancies and the wage offers accompanying these vacancies. Overall, the evidence points to possible long-term shortage conditions based on the wage-employment results, but the lack of complementary evidence on nonwage adjustments suggests viewing this conclusion with appropriate caution.

Table 10.4
Employment by Selected Occupation, 2002 and Projected 2012

	Employment		Change	
Occupation	**2002**	**2012**	**Number**	**Percentage**
Total, all workers	144,014,600	165,318,670	21,305,070	14.8
Industrial engineers	158,252	175,103	16,851	10.7
Mechanical engineers	215,103	225,427	10,323	4.8

Summary

This paper has attempted to provide a road map on how to assess whether long-term labor shortage conditions exist for occupations. It was argued that modeling separate labor supply and demand functions is too problematic, especially given the challenges of modeling other forms of adjustment that employers make when faced with difficulties in hiring workers. An alternative, equilibrium-based approach was adopted, in which occupations are initially screened on the basis of whether there have been significant increases in nominal wages accompanied by relatively weak employment responses. A second screen to test for the possibility that the weak employment response is the result of occupations being highly concentrated in declining industries was administered. Finally, a host of other adjustment variables were examined in an attempt to detect nonwage adjustments that firms may pursue to find workers.

This paper does not claim to provide the "perfect" road map for the determination of long-term shortage conditions. Indeed, the triage approach adopted screens occupations on the basis of characteristics that, while consistent with the classical view of shortages, are not always necessary to produce significant difficulties in finding workers. Owing to data gaps, one natural characteristic that was not examined is the presence of persistent occupational vacancies despite rising wage offers. In the end, it is the preponderance of evidence, along with an admittedly conservative approach, that filters for and identifies the occupations most likely to be experiencing long-term shortage conditions.

CHAPTER ELEVEN

U.S. Census Bureau Data and the Science and Technology Workforce

Comments by Robert Kominski[1]

Although the U.S. Census Bureau gathers extensive data on the entire U.S. population, its role in helping to define the shape of and to track the country's science and technology workforce is actually quite limited. Generally, the science and technology workforce is considered to be made up almost exclusively of those with at least one postsecondary degree. It is important to note that, within the entire U.S. population, about 1.2 percent hold PhDs and only roughly a quarter hold a postsecondary degree of any kind. For this reason the Census Bureau has generally not attempted to address issues of measuring the science and technology workforce in detail. However, some of the data collected through the bureau's ongoing programs—such as the Current Population Survey, the Survey of Income and Program Participation, the decennial census long form, and the American Community Survey—may be relevant to the discussion, provided the limitations of such data are recognized.

Perhaps the most relevant single piece of data collected by the bureau pertains to the educational attainment of individuals. The Census has collected some piece of information on this topic for decades. Before 1940, educational attainment was measured simply by asking whether a person could read and write—i.e., if they were literate. Beginning in 1940, the Census asked for the highest grade level completed by each individual in the household. In 1990, this question was expanded to include categories for those completing postsecondary schooling beyond college. This new question brings us closer to capturing the educational attainment of the U.S. population, but still missing is crucial information about the knowledge, skills, and abilities of the technical workforce.

Inherent in the current educational attainment question are several limitations and exclusions. Since the form asks for the highest grade level completed, respondents are asked to report only on the extent of their formal education. Not measured are achievements and training received through nontraditional means. Similarly, work experience and relevant vocational or on-the-job training are not reflected in the respondents' answers. Detailed characteristics of the education received that would further define the actual shape of the science and technical workforce, such as the field of concentration, quality of training, or type of institution, also remain unknown.

There are more-general exclusions in the Census's measurement of educational attainment that are especially relevant when measuring the science and technology workforce. Also significant are the growing number of people obtaining training "certification" by

[1] Robert Kominski is Assistant Chief for Social and Demographic Statistics in the Population Division of the U.S. Census Bureau, Washington, D.C. Transcribed from Mr. Kominski's overhead slides and notes by Gabrielle Bloom, a research assistant with the RAND Corporation, Arlington, Virginia.

completing targeted programs usually meant to teach a very particular skill set (e.g., fluency in a particular computer program or software suite). This is the so-called parallel postsecondary universe (Adelman, 2000). The Federal Interagency Committee on the Measurement of Educational Attainment, consisting of representatives from various federal statistical agencies, has identified the difficulty in tracking, counting, and quantifying the number and skill level of those completing such programs as the current number one barrier to measuring educational attainment. In one count of the most-popular information technology certification programs, the number of certificates granted as of the year 2000 was close to 2 million (Adelman, 2000). It is unclear what proportion of these certificate recipients are being excluded or miscounted using current methods of measuring educational attainment. Since the majority of these programs offer development of technical skills, this problem is particularly relevant in the effort to measure the science and technology workforce.

The data collected by the Census on occupation can also be useful, but, again, there are limitations and a potential for error that must be understood. The decennial census question that is currently used to obtain information on current employment appears on the long form and asks respondents to describe the kind of occupation and their most important duties or activities. These answers are then collapsed into 509 detailed codes. Although these codes are highly "detailed," it is difficult to be sure that the correct occupation or meaning of job description is captured, as these definitions change over time and some data sets may require collapsing because of small sample sizes. These discrepancies are compounded by the lack of data on employment history, which would be useful in and of themselves as a means of understanding the paths taken by participants in the science and technology workforce. The use of occupational data as an indicator of workforce skill or competency level is also difficult because jobs do not necessarily correspond with standardized training.

In addition to the limited data on educational attainment and current occupation, the Census Bureau collects potentially useful data in some of its special studies, such as the Bureau of Labor Statistics supplements to the Current Population Survey, or from surveys sponsored by other agencies that focus on labor force issues. The Census Bureau generally does not analyze these datasets but does make the data publicly available. Also of potential relevance to this issue are a few special data collections on computer and Internet use.[2]

The various types of data described above have the potential to contribute to the measurement of the U.S. science and technology workforce. The real strength of Census Bureau data, however, lies in their ability to detail relationships among data items, as well as providing geographic specificity. These data can be used to create detailed portraits of specific subpopulations or small pieces of geography. When used in combination with data defining the science and technology workforce subpopulation, such variables as age, race, sex, language, and disability status can be as important as more obvious variables, such as education or occupation. Such information can help to describe more accurately any economic, racial, or social disparities that may exist in the current workforce and aid in the development of more targeted solutions for the future. Similarly, this type of "base population" data can also be used to drive population projections that may be of use when deciding on long-term efforts.

[2] These collections are known as the computer use supplements to the CPS and exist for the years 1984, 1989, 1993, 1997, 1998, 2000, 2001, and 2003.

Definitional discrepancies and technical limitations make data items in the recurring surveys issued by the U.S. Census Bureau of limited use when it comes to measuring the nation's science and technology workforce. The data are not detailed enough in scope to provide a thorough accounting of the knowledge, skills, and abilities of the workforce. The power of the Census Bureau data, however, comes from their ability to place segments of the population in the context of many other social, economic, and demographic characteristics. Significant differences in the current status and past trends of characteristics within a subpopulation can exist across geographical regions or socioeconomic groups. The possible applications of such analyses in the effort to measure and predict flows in the U.S. science and technology workforce have not been fully explored.

Reference

Adelman, C., *A Parallel Postsecondary Universe: The Certification System in Information Technology*, Jessup, Md.: Department of Education, Education Publications Center, 2000. Online at http://www.ed.gov/pubs/ParallelUniverse/title.html (as of January 6, 2004).

CHAPTER TWELVE

Opportunities and Challenges at the National Center for Education Statistics

C. Dennis Carroll[1]

Abstract

The National Center for Education Statistics (NCES) program at the postsecondary education level provides statistical information used by planners, policymakers, and educators in addressing a multitude of issues. With such studies as the annual Integrated Postsecondary Education Data System and the National Postsecondary Student Aid Study, some high-quality data on education related to science and technology workers are available. Users will find friendly tools to access and analyze these data, in real time, over the Web.

Introduction

The NCES program at the postsecondary education level provides statistical information used by planners, policymakers, and educators in addressing a multitude of issues. The initial source of most information is the annual Integrated Postsecondary Education Data System (IPEDS). It provides a variety of data on the nation's public and private postsecondary institutions. Special studies of student financial aid, postsecondary faculty, and various longitudinal studies complement the basic data in IPEDS and use IPEDS as the sampling frame.

One of the largest postsecondary sample studies is the National Postsecondary Student Aid Study (NPSAS). This is a comprehensive nationwide study designed to determine how students and their families pay for postsecondary education and to describe some demographic and other characteristics of those enrolled. The study is based on a nationally representative sample of students in postsecondary education institutions, including undergraduate, graduate, and first-professional students. Students attending all types and levels of institutions are represented, including public and private not-for-profit and for-profit institutions and less-than-2-year institutions, community colleges, and 4-year colleges and universities. The first NPSAS study was conducted during the 1986–87 school year; subsequent studies have been carried out during 1989–90, 1992–93, 1995–96, and 1999–00. Each of the NPSAS surveys provides information on the cost of postsecondary education, the distribution of financial aid, and the characteristics of both aided and nonaided students and their

[1] Charles Dennis Carroll is Associate Commissioner, National Center for Education Statistics, Postsecondary Studies Division, Washington, D.C.

families. Following each survey, NCES publishes three major reports: Undergraduate Financing of Postsecondary Education, Student Financing of Graduate and Professional Education, and Profile of Undergraduates in U.S. Postsecondary Education Institutions. The latest NPSAS full-scale study is under way for the 2003–04 academic year.

Beginning Postsecondary Students (BPS) studies follow students at the beginning of their postsecondary education. Initially, students in the NPSAS surveys are identified as being first-time beginners of undergraduate studies. These students are asked questions about their experiences during, and transitions through, postsecondary education and into the labor force, as well as family formation. Transfers, persisters, stopouts and dropouts, and vocational completers are among those included in the studies. In the first BPS study, students were identified in the 1990 NPSAS as being first-time beginning postsecondary students during the academic year 1989–90. These students were followed in 1992 and in 1994. A second cohort of first-time beginning students was identified in the 1996 NPSAS, with follow-ups performed in 1998 and in 2001. The next cohort will be identified in the 2004 NPSAS.

Baccalaureate and Beyond (B&B) studies follow students who complete their baccalaureate degrees. Initially, students in the NPSAS surveys are identified as being in their last year of undergraduate studies. Students are asked questions about their future employment and education expectations, as well as about their undergraduate education. In later follow-ups, students are asked questions about their job search activities, education, and employment experiences after graduation. Individuals who had shown an interest in becoming teachers are asked additional questions about their pursuit of teaching and, if they are teaching, about their current teaching position. In the first B&B study, students were identified in the 1993 NPSAS who had completed their degree in the 1992–93 academic year. These students comprised the first B&B cohort and were followed up in 1994, 1997, and 2003. A new B&B cohort began with the 2000 NPSAS and involved only a 1-year follow-up in 2001.

The National Study of Postsecondary Faculty (NSOPF) includes a nationally representative sample of full- and part-time faculty and instructional staff at public and private not-for-profit 2- and 4-year institutions in the United States. The study was initially conducted during the 1987–88 academic year and was repeated in 1992–93 and 1998–99. The latest data collection is under way for 2003–04.

Finally, there are two longitudinal studies of high school students. The National Education Longitudinal Study (NELS), which began with an 8th grade cohort in 1988, provides trend data about critical transitions experienced by young people as they attend school and embark on their careers. Data were collected from students and their parents, teachers, and high school principals and from existing school records, such as high school and college transcripts. Cognitive tests (math, science, reading, and history) were administered during the base year (1988), first follow-up (1990), and second follow-up (1992). The third follow-up (1994) and fourth follow-up (2000) used computer-assisted telephone interviews. The 2000 follow-up included collection of postsecondary transcripts.

The Education Longitudinal Study of 2002 (ELS) is a survey that will monitor the transitions of a national sample of young people as they progress from 10th grade to, eventually, the world of work. ELS will obtain information not just from students and their school records, but also from students' parents, teachers, librarians, and school administrators.

Coding Program and Major Field of Study

Within the IPEDS data, there are data describing the numbers of degrees and other awards granted by program of study. These completion data are based on institution reports using the Classification of Instructional Programs (CIP). The CIP is a taxonomic coding scheme that contains titles and descriptions of instructional programs, primarily at the postsecondary level. CIP was originally developed to facilitate NCES's collection and reporting of postsecondary degree completions, by major field of study, using standard classifications that capture the majority of program activity. The CIP 2000 edition is the third revision of the CIP. It was originally published in 1980 and was revised in 1985 and again in 1990. The CIP is now the accepted federal government statistical standard for classifying instructional programs; in addition, the 2000 edition has been adopted by Statistics Canada as its standard taxonomy for field of study.

The level of detail within the CIP is displayed within the listing for engineering. Unfortunately, this level of detail is beyond the capacity of the NCES sample surveys. Within NPSAS, for example, major field of study is coded using a limited number of the 6-digit CIP codes.

Science and Technology Data Within NCES Studies

Within the IPEDS, the only science and technology data are from the completion component, describing counts of degrees and other awards by CIP codes, gender, and race and ethnicity. There are no data describing enrollments, finance, graduation rates, or faculty and staff.

Within the sample surveys (e.g., NPSAS), the major field-of-study variables are part of a larger microrecord data set that allows substantially more detailed tabulation with other variables. These data sets include household composition, goals and aspirations, plans, employment and labor force measures, educational experiences, and financial data in addition to gender, race, and ethnicity. Unfortunately, about 18 percent of the cases have no declared major.

Analyses of these NCES data are substantially easier today with the availability of Web tools that allow users to tabulate and access files online. All of the postsecondary sample surveys have online Data Analysis Systems (DAS) that are fourth generation, full featured, and relatively user friendly (see NCES, undated a). The IPEDS data may be accessed online (see NCES, undated b), and a DAS should be available in 2004.

The Good News

There is a substantial amount of good news concerning NCES data. Data quality has never been better. The data collections incorporate Web systems and computer-assisted telephone protocols that include editing and consistency checks in real-time. NCES is past the initial set of collection problems experienced by one-time surveys and the cyclical collections have matured with higher quality over time. Finally, the NCES statistical standards (see NCES, 2002) have improved operations to produce higher-quality data.

There is more and better training for data collectors, analysts, and (for IPEDS) respondents. Online tutorials and face-time training are available. Well-trained data collectors (and respondents) should produce better data.

NCES is quicker. For example, IPEDS data for academic year 1995–96 were available in 1999; in comparison, IPEDS data for academic year 2001–2002 were available in early 2003. We are trying to speed up the processes for the sample surveys as well.

The Bad News

Institutional and individual nonresponse rates have been increasing. For all the effort focused on nonresponse, this is clearly bad news. Some of the approaches used by NCES to deal with this problem have been incentive payments to individuals, reimbursements and honoraria to institutional staff, and the identification and deployment of nonrespondent conversion specialists.

Timeliness remains a problem. NCES Customer Service surveys routinely identify timeliness as a major concern. There are three sources of timing problems. First, some data are systemically a challenge. The IPEDS data on institutional finance are derived from audited, general-purpose financial statements. As such, the data cannot be collected until the accountants complete their audits. This allows collection 8 months after the close of the academic year. Second, dealing with nonresponse takes time. Conversion efforts usually require 3 to 6 months. Third, NCES review procedures, including several iterations of inspection, require 6 to 8 months. Improving timeliness is one of the most difficult problems.

Gaps in the data are widening. At the institutional level, part-time enrollment, co-enrollment, distance education enrollment, noncredit enrollment, and stopout persistence patterns are more frequent. All of these are difficult to capture with existing data-collection systems. As NCES has watched distance education evolve, for example, each set of items used for data collection has been out of date within two years.

Finally, resources are limited. The NCES staff count has declined, and some staff are nearing retirement.

Summary

NCES has postsecondary data collection systems in place to provide some high-quality data on education related to science and technology workers. Users should find friendly tools to access and analyze these data, in real time, over the Web. However, users are cautioned to observe response rates and the age of the data.

References

National Center for Education Statics, Data Analysis System, undated a. Online at http://nces.ed.gov/dasol/index.asp (as of June 4, 2004).

_____, IPEDS Peer Analysis System, undated b. Online at http://nces.ed.gov/ipedspas/ (as of June 4, 2004).

_____, Statistical Standards, 2002. Online at http://nces.ed.gov/statprog/stat_standards.asp (as of June 4, 2004).

Part III
Summary and Recommendations

CHAPTER THIRTEEN

Rapporteur's Summary

William P. Butz[1]

This conference afforded an extraordinarily rich cross talk among experts who make policy that affects the STEM workforce, analysts who study it, and statisticians who design and collect the data that support both policy and analysis. Despite the diverse perspectives represented here, there was an underlying consensus on the purpose of these deliberations: to identify specific actionable improvements in the data system for decisionmaking.

Keeping that aim in view, this section summarizes the conference's dominant themes. It first proposes that the questions addressed can effectively be distilled to three: Who are the decisionmakers that need the data? What data do they need? What data can suppliers provide? In addition, a set of crosscutting themes relevant to more than one of these questions emerged.

Decisionmakers in the Markets for STEM Workers

Who are the decisionmakers to be informed? Conferees identified five groups of decisionmakers. In addition, the discussion pointed to a sixth group with special needs: researchers.

Policymakers in Science Funding Agencies

These agencies, NSF and NIH principally among them, strongly influence the numbers of STEM students in the educational pipelines. At the same time, they provide much of the support in many fields for the job positions and laboratories that will or will not await these students when they enter the job market. When deciding, for example, whether to increase or decrease fellowship and postdoctoral support, policymakers should balance trend information on numbers of positions by field against data on numbers of students becoming available to fill them.

Program and Human Resource Administrators in Federal and Other Public Agencies That Employ Scientists and Engineers

These officials infer STEM personnel requirements from prospective changes in their agencies' responsibilities and in the science that underlies their work. In satisfying these requirements, they must balance uncertain staff retirement and retraining prospects against the possibility of hiring replacements and new types of specialists from the labor market. Program and budget planning typically requires human resources departments to specify these hiring

[1] William P. Butz is President and Chief Executive Officer of the Population Reference Bureau in Washington, D.C., and a consultant to RAND.

plans several years ahead of the need. Some agencies mitigate some of the risk involved by contracting work requiring STEM personnel to the private sector.

Private-Sector Employers, Colleges, and Universities

These entities face similar balancing acts as they forecast the demand for their products and services, the staff on hand to produce them, and the availability of suitable personnel to fill replacement and new positions. The increasing possibility of outsourcing work to other countries does not diminish the requirement for planning and balancing; to the contrary, it extends the geographic domain for which data on the STEM workforce are valuable.

Members of the STEM Workforce

Whether driven by scientific discovery, technical development, market demand for the products of technology, or federal funding, wide swings in employer demand for successive science and engineering specialties have characterized U.S. labor markets since the mid-1960s. Most American STEM workers now undergo some form of mid- or late-career retraining—in colleges, employers' laboratories, or on the fly. The more accurately workers can discern the changing demand for their specialty or for other specialties not far removed, as well as the changing supply of workers in these fields, the better they can identify approaching job trouble or opportunities. Families, employers, and the nation will suffer less economic deprivation, production cost increases, and unemployment, respectively.

Students and Their Advisors

A high school senior working toward a technical associate degree only needs to look ahead two or three years to assess what jobs will be available to reward that training. A high school senior working toward a doctorate and a postdoctoral appointment, however, must try to look ahead many more years—at least a dozen, on average, in biology, perhaps a year or two less in other sciences. College seniors must also look ahead several years in deciding among careers in science, on the one hand, or in medicine, dentistry, veterinary science, law, or business, on the other. Thus, long-term information about job markets, trends, and prospects can be invaluable to career decisions.

While no crystal ball is available, current information about trends in job offerings by degree field, the number of students already in the pipeline, and underlying changes in government funding can at least suggest which fields have unattractive near-term employment prospects.

Labor Market Researchers

The estimates and insights of the economists, sociologists, and psychologists who study the labor market help others make informed decisions. This work clarifies the influences on student choice of college major and graduate field, on worker decisions about job choice and retirement, and on employer decisions about job characteristics and recruitment. Such research requires data beyond counts of workers and jobs.

At every level—from science policy administrators to guidance counselors to high school students—there are decisionmakers. When federal statistical agencies or industry associations do not provide objective and timely data on STEM jobs, job holders, and job seekers, responsible decisionmakers will instead collect anecdotes, impressions, and whatever else they can find to inform one of life's most consequential decisions. Moreover, when these responsible organizations do not provide credible analyses of these data—analyses that turn the numbers into useful information—decisionmakers perform their own analyses, however implicit and simplistic.

Crosscutting Themes

The conference discussions yielded a set of specific data requests from the researchers and decisionmakers represented, as well as responses from the statistical agencies represented, about priorities and practicalities. Cutting across these specific data items were some general themes and issues that spanned the different decisionmaking and data domains and connected many of the otherwise diverse topics.

The first of these—general STEM competency—is ideally a national goal that underpins maintaining a quality STEM workforce. The next three themes—revisiting assumptions about STEM careers, data coordination and collection, and identifying workforce shortages and surpluses—are means of achieving a range of goals. The final theme, the global dimension, is an emerging variable that must be factored into STEM workforce decisionmaking.

A Broader Context: General STEM Competency

It is important to bear broader contexts in mind when considering STEM workforce data issues. Among the most important of these is promoting the goal of STEM competency among the general public. STEM competency is important at all levels from the "ordinary citizen" through those holding science doctorates. Information about the effectiveness of K–12 science and mathematics education is important because it both lays the foundation for entering and succeeding in scientific fields and contributes significantly to the success of informed citizens in any occupation.

Revisiting Key Assumptions About Training and Careers

Data collection and analysis are often influenced by assumptions about typical career paths and job progressions, which may not accurately reflect shifting and complex occupational realities. Many of these assumptions need to be revised. For example,

- Industry employment is quantitatively important in most STEM fields. While this may seem evident, some data collection and analyses still proceed as if newly minted scientists move solely or mainly into college teaching and as if professors do not commonly work in both worlds, simultaneously or sequentially.
- Life-cycle progressions—starting with choice of college major, moving through career advancements and choices, and ending with retirement—are poorly understood for people in STEM professions. Even for the archetypal career path into college teaching, we lack basic descriptions of these progressions, the key decision points, and the influencing factors.

Data Coordination and Collection Issues

The conference discussion focused primarily on federal data collection and statistical analysis. This point raised several issues:

- Further coordination of statistical concepts and definitions and productive sharing and linking of data across federal statistical agencies would contribute substantially to the completeness and usefulness of data for STEM decisionmakers of every type.
- Although the data that federal statistical agencies collect command by far the most attention from researchers and STEM decisionmakers, considerable data are also available from federal and state administrative agencies and from professional associations and other private sources. For some important concepts, the latter organizations are the only current sources. Coordinating and documenting these diverse collections of data, presumably by the federal statistical system, could provide a considerable payoff.
- Expanding the access nonfederal researchers have to data would also increase the useful information available to all STEM decisionmakers. Directed research and coordinated federal effort could mitigate the effects of confidentiality provisions and the technical and bureaucratic difficulties of linking related data sets.
- When possible, it would be best to satisfy the requirement for new or different data by adding question modules to existing survey programs, rather than beginning new surveys.

Identifying Workforce Shortages and Surpluses

Existing data suitable for monitoring possible shortages or surpluses among STEM workers by field are not sufficient for understanding and improving observed patterns:

- People trained in STEM fields but working in other occupations must be accounted for to document and understand the existence of shortages or surpluses in STEM fields. The progression of scientists to general administration is only one example.
- Cross-sectional and longitudinal data are both critical. In particular, labor-market research requires data on worker and employer characteristics and on the community context, as well as a longitudinal survey structure that documents life-cycle changes for the same samples of workers. Such data are essential for estimating the labor supply elasticities that characterize student and worker responsiveness to changing job availability and earnings.
- In addition, quick-turnaround data systems can and should be designed to identify STEM fields at risk of emerging shortage or surplus. These relatively crude indicators of job openings, job applications, salaries, graduate student enrollments, and job placements, for example, would direct more substantial ongoing data collections toward particular STEM fields, revealing whether the concern is justified. In the meantime, students contemplating fields of study would have tentative indicators several years earlier than they do at present, a substantial difference at that life stage.
- Data on workers are many times more plentiful than data on their employment conditions. Expending the time and money necessary to document the demand side of the scientific labor market—job slots, job offers, characteristics of facilities and equipment that complement or substitute for labor input, characteristics of demand for the products or services the employer produces—will have a big payoff in elucidating the sources of STEM worker shortages and surpluses.
- Additional data on how firms adjust to occupational labor shortages are critical to understanding whether shortages reflect a short-term lack of equilibrium or long-

term structural shortages. To the extent that rising wages do not elicit a sufficient supply of workers in the short, medium, or even long runs, firms may adjust in other ways, often affecting working conditions and the quality of STEM labor. These adjustments can include changes in overtime hours, increased use of immigrants, lowering educational attainment requirements, or the use of offshore or contract labor.
- Finally, forecasting is an important issue. Forecasts of labor market conditions received little attention and emphasis at the conference because most participants considered them too inaccurate to be useful. However, if such forecasts could be made more accurate, they would be valuable.

The Global Dimension

A new variable that must be taken into account in considering these is the global dimension. Changes in the numbers of foreign students enrolled here and the foreign scientists employed here; in the international movement of U.S. scientists during the course of their work; and, most recently, in the outsourcing of technical and even scientific work to less-expensive workers abroad—these trends have integrated U.S. scientific institutions and their employees with foreign counterparts, in many cases quite inextricably. Decisions based solely on U.S. data about Americans may increasingly paint an incomplete and possibly misleading portrait of the market for STEM workers.

What Data Do Decisionmakers Require? What Data Can Producers Supply?

Workshop participants from the research community and from organizations that employ STEM workers identified particular data items and data changes that would improve analysis and decisionmaking. In turn, participants from the federal statistical system addressed each of these requirements, pointing out some that are already being satisfied, others that could be met with more or less difficulty, and still others that might be met by substituting available proxy information. Indeed, an immediate payoff of the workshop was that data users discovered that some requested data are already available, either as described or in the form of a useful proxy. Below are descriptions of the data that the users—the organizations and researchers—requested, alongside statistical agency indications of current and potential availability. The workshop participants' presentations in Part II provide the background justification and context for most of these requests.

- Education and training
  - Collect median and variance of length of graduate program and probability of completion data, by field. Not available apart from statistics kept by several professional associations.
  - Collect data on numbers and characteristics of persons who drop out of a STEM educational process, by discipline, or who leave STEM occupations while still employed, by occupation entered. Not available or planned.
  - Collect data on numbers and characteristics of "false starts" in the educational process, by level and field. Not available or planned.

- Compile clearinghouse of demonstrations and experiments designed to improve training, retraining, job search, or career advancement of STEM workers. Not available or planned.
- Compile information on alternative educational paths to STEM careers. Not available or planned.
- Collect data following foreign students through the U.S. educational process. Student and Exchange Visitor Information System (SEVIS) data, still being developed, are a big improvement.
- Define an "understanding of science" in the citizenry and articulate how this is measured. Not available or planned.
- Collect data on numbers of students by discipline, degree program, stage, and citizenship status, reported within three months of the beginning of the academic year. NCES and SRS collect a subset of these data.

• Degree holders and postdoctorates
  - Collect data on numbers and characteristics of persons holding postdoctoral positions. SRS is beginning methodological research on administrative data that NIH and NSF have collected.
  - Collect data on holders of multidisciplinary degrees comparable to data available for traditional disciplines. SRS is conducting methodological research.
  - Study numbers and characteristics of foreign STEM degree holders in the United States. Not available or planned.
  - Compile data on STEM workers with degrees and certifications below the baccalaureate level—for example, technical associate in arts degrees and software-manufacturer training certifications. SRS and NCES are working collaboratively to develop these data.
  - Add STEM workers with clinical degrees and without doctorates to the Survey of Earned Doctorates (SED) and the Survey of Doctorate Recipients (SDR). Not available or planned.
  - Include doctorates from foreign institutions in the sample frames for federal surveys of STEM workers. Not available or planned.

• STEM workforce: jobs and occupations
  - Collect employment data that provide greater detail on occupations at local, state, and national levels. Available at BLS.
  - Collect data on job vacancies and turnovers, by occupation. Not available or planned.
  - Compile data on newly emerging occupations. Not available or planned.
  - Compile detailed data on net job replacement requirements, by occupation. Available at BLS.
  - Study job competencies that employers seek, apart from degrees and certifications. Not available or planned.
  - Compile data on hiring at venture capital firms, by occupation, as a leading indicator of needed skills. Not available or planned.
  - Compile data on the number of job offers, by occupation. Not available or planned.

  - Follow up on the 30 percent of SED respondents who report "no definite plans." Not available or planned.
  - Track STEM workers in industry employment. Not available or planned.
  - Create longitudinally matched data on employers and STEM employees in the U.S. Census Longitudinal Establishment Data (LED) file. This should not be limited to the Census LED but should be broadened to include linked employer-employee establishment data files. Not available or planned.
  - Assemble characteristic data on large samples of STEM workers more frequently. The Census Bureau's American Community Survey will produce such data on very large samples between censuses.
  - Collect data on job offers, acceptances, and salaries in aggregated occupational categories, reported three months after close of the survey period. Not available or planned.

- Metrics and Databases
  - Create a national STEM database. Not available or planned.
  - Develop a common set of concepts and definitions across federal statistical agencies. Talks among agencies are continuing.
  - Develop measures of lifetime risk of unemployment and career change, by STEM discipline. Not available or planned.
  - Create a unified bibliography of publications and citations regarding the STEM workforce. SRS is working on methodological improvements to existing bibliographies.
  - Maintain and improve response rates for STEM surveys. Statistical agencies maintain substantial research on this challenge.
  - Facilitate data-sharing across federal statistical agencies to produce matched files for analysis. OMB and the statistical agencies are working on this challenge.
  - Make it easier for university researchers to use federal microdata. OMB and the statistical agencies work continually on this challenge.
  - Construct an omnibus sample frame and a related program of "snap" surveys to produce very quick turnaround data on particular policy-related questions. Not available or planned.

- International issues. Nearly all the other recommendations have an international dimension that needs to be documented. We therefore list the international data issues separately to highlight their importance, even though they relate closely to many of the other items:
  - Assemble internationally comparable data on numbers of STEM workers and of STEM jobs. Not available. NCES is working with Statistics Canada, OECD, and UNESCO to develop comparable concepts and definitions. NCES is working with the Department of Homeland Security (DHS) to obtain information on degrees and occupations of STEM visa holders working in the United States.
  - Collect information on international mobility of foreign STEM workers educated in the United States. SRS is conducting methodological research based on DHS data.
  - Collect information on STEM jobs outsourced to other countries by U.S. employers. Not available or planned.

– Survey employers on why and under what circumstances they move jobs overseas. Not available or planned.

Concluding Observations

Again and again, the conference discussion returned to the connection between data and decisionmaking and reiterated the basic point that decisionmaking does not grind to a halt in the absence of adequate data. It simply proceeds with inadequate data. Employers and managers who lack a credible information base produced by statistical experts may base decisions on information and analyses that they themselves have produced, often on the fly, or that are produced by others lacking statistical expertise. Administrators of science funding agencies who lack such information may base funding allocations across scientific disciplines on judgments about where the science is most exciting or where other support is lacking, to the detriment of students encouraged toward fields, however exciting, without waiting jobs. Moreover, without understanding key decision points for STEM students and workers, universities and their science funders cannot efficiently design interventions to affect such decisions. Among all these decisionmakers, students and workers are the most disadvantaged, for they typically command insufficient resources to uncover any but the most rudimentary information about trends in potentially interesting fields. And yet, ironically, it is they who bear the largest burden of mistaken decisions—lengthy training and uncertain outcomes, job insecurity, and potential disillusionment with the scientific enterprise.

Of the numerous data requirements that surfaced at the conference, the eight highest priority needs are presented in detail in Chapter Fourteen.

CHAPTER FOURTEEN

Priority Data Improvements

This chapter presents RAND's assessment of which data-improvement needs should be given priority. The assessment is based on all the presentations and discussion at the conference, as well as on our own insights and experience. We have considered these themes and data requests in the broad context of technical difficulty, resource scarcity in the federal statistical agencies, and potential for contributing to the broadest range of decisionmakers and researchers. The following eight data improvement priorities are the result.

Current Job Market Conditions

Data that do not become available until two or three years after the fact, with analyses still to follow, are of little value for important decisionmaking purposes. Students making career decisions must often base them on obsolete information in rapidly changing times. Likewise, federal science-funding agencies are far out of date when deciding which fields require increased (or decreased) fellowship and postdoctoral support to prevent shortages (or surpluses) several years hence and which fields require increased (or decreased) grant support to provide jobs for students already in the pipeline. In this area, the following are most important:

- Data on job offers, acceptances, and salaries in aggregated occupational categories should be reported three months after the end of the survey. To serve as models, unemployment statistics and final national account statistics are reported one week and one month, respectively, after the end of the survey.
- Data on numbers of students by discipline, degree program, stage, and citizenship status should be reported within three months of the beginning of the academic year.

Make these data readily available to students and student advisors within the three-month window.

Comparative Graduate Program Data

Students currently have only an anecdotal basis for estimating and comparing across universities and fields of study the largest cost of career preparation: the amount of time in graduate study. In this area, the following is most important:

- Data must be available on median and variance of length of graduate study and the probability of completion, sorted by field.

These comparable data, when widely available, might also focus universities' and professional associations' attention on this challenge.

Private Industry Data

Industry now hires almost 40 percent of U.S. STEM doctoral graduates. Ignoring these movements between academe and industry seems as unlikely to produce accurate identification and understanding of STEM worker shortages and surpluses as does ignoring international movements. In this area, the following is most important:

- The sample frames of relevant surveys of STEM workers should include private-industry employers and their workers.

STEM Career Paths

The number of students who begin a STEM academic track but change to another is unknown but is thought to be large. The number of STEM degree holders engaged in non-STEM occupations is likewise unknown but seems large. These "leakages" likely run mostly in one direction: out of STEM careers. While such movements are not necessarily undesirable for the students or society, they are probably the largest component of STEM personnel loss in the system, next to retirement. Knowing virtually nothing about these people—their numbers, characteristics, or circumstances of change—leaves efforts to identify impending shortages or surpluses vulnerable. Efforts to influence these career decisions are similarly hobbled. In this area, the following are most important:

- Data on the numbers and characteristics of persons who drop out of a STEM educational process, by discipline, or who leave STEM occupations while still employed, by occupation entered, are important to document.
- Follow-up data on the 30 percent of SED respondents who report "no definite plans" should be collected and reported.

Global Workforce

Global aspects of STEM education and employment in the United States have become so quantitatively large and systemically integrated that questions of STEM worker shortages or surpluses can no longer be meaningfully answered without corresponding global data:

- Numbers and characteristics of foreign STEM students and workers by discipline in the United States must be documented, including incorporation of doctorates from foreign institutions in sample survey frames.
- Information on STEM jobs that U.S. employers have outsourced to other countries must be collected; the primary source of these data may be a federal statistical agency not normally engaged in collecting STEM data.

International comparability of concepts, definitions, and formats, while important for other purposes, is less directly important than the above for decisionmakers in the United States—employers, students, and science funders.

Moving Forward

As statistical agency administrators discussed at the conference, several of these priority requirements are under discussion, and more are well in planning. Others, though, are not on the drawing board. Some will require additions to sample frames, others additions to questionnaire content, still others entirely new data collection programs. Several may require new collaborations across agencies and international borders. To the extent that identifying and understanding shortages and surpluses of STEM workers are important to U.S. decisionmakers—that is, to students, workers, employers, and science funders—these priority data additions and improvements are correspondingly important.

APPENDIX A

Conference Agenda

How Will We Recognize a Shortage or Surplus in the Scientific and Technical Workforce? Improving the Data System for Decisionmaking[1]

December 11, 2003

Morning	"Does America Face a Shortage of Scientists and Engineers?
9:30–10:00	Welcome and Introduction: Kathie L. Olsen, OSTP; Ralph Gomory, Sloan Foundation
10:00–10:30	"Do We Need More Scientists?" Michael Teitelbaum, Sloan Foundation
10:30–11:00	Discussants: Michael P. Crosby, NSF; Ronald Ehrenberg, Cornell
11:00–11:15	Break
11:15–12:00	Participant discussion
12:00–1:00	Lunch
Afternoon	"What Data Do Labor Market Researchers Need?"
1:00–1:30	Richard Freeman, Harvard and NBER
1:30–1:45	Paula Stephan, Georgia State University
1:45–2:15	Participant discussion
2:15–2:30	Break
	"What Data Do Science and Technical Policymakers Need?"
2:30–2:45	Judith Ramaley, NSF
2:45–3:00	Walter Schaffer, NIH
3:00–3:15	Patrick Simpkins, NASA
3:15–3:30	Ben Wu, Deputy Undersecretary of Commerce for Technology
3:30–4:15	Participant discussion
4:15–4:30	Rapporteur's summary of needed improvements in the data system

December 12, 2003

Morning	"Meeting the Data Needs: Opportunities and Challenges in the US Government"
9:00–9:15	Opening remarks by panel chair, Katherine Wallman, OMB
9:15–9:45	"Opportunities and Challenges in the National Science Foundation" Lynda Carlson, NSF
9:45–10:15	"Opportunities and Challenges at the Bureau of Labor Statistics" Michael Horrigan, BLS
10:15–10:30	Break
10:30–11:00	"Opportunities and Challenges at the Census Bureau" Robert Kominski, Census Bureau
11:00–11:30	"Opportunities and Challenges at the National Center for Education Statistics" C. Dennis Carroll, NCES

[1] Held at the RAND Corporation, December 11–12, 2003.

11:30–12:15	Participant discussion
12:15–1:30	Lunch
Afternoon	"A Program for Data Improvement"
1:30–2:00	Rapporteur's summary of opportunities and challenges for meeting the data needs
2:00–3:30	Participant discussion
3:30–3:45	Conference wrap-up and next steps

APPENDIX B

Biographical Notes on Contributors

Presenters

Lynda T. Carlson is Director, Division of Science Resources Statistics, National Science Foundation, Arlington, Virginia.

Charles Dennis Carroll is Associate Commissioner, National Center for Education Statistics, Postsecondary Studies Division, Washington, D.C.

Michael P. Crosby is an Executive Officer of the National Science Board, Arlington, Virginia.

Ronald G. Ehrenberg is the Irving M. Ives Professor of Industrial and Labor Relations and Economics at Cornell University in Ithaca, New York, and a research associate at the National Bureau of Economic Research, Washington, D.C.

Richard B. Freeman is Ascherman Professor of Economics at Harvard University in Cambridge, Massachusetts, Co-Director of the Labor and Worklife Program at the Harvard Law School, and Director of the Labor Studies Program at the National Bureau of Economic Research, Washington, D.C. He is also Co-Director of the Centre for Economic Performance at the London School of Economics and Visiting Professor at the London School of Economics.

Michael W. Horrigan is Assistant Commissioner in the Office of Occupational Statistics and Employment Projections at the Bureau of Labor Statistics, Washington, D.C.

Robert Kominski is Assistant Chief for Social and Demographic Statistics in the Population Division of the U.S. Census Bureau, Washington, D.C.

Kathie L. Olsen is Associate Director with the Office of Science and Technology Policy in the Executive Office of the President, Washington, D.C.

Jean Pomeroy is a senior policy analyst with the National Science Board, Arlington, Virginia.

Judith A. Ramaley is Assistant Director, Education and Human Resources Directorate at the National Science Foundation in Arlington, Virginia. She also holds a presidential professorship in biomedical sciences at the University of Maine-Orono and is a Fellow of the Margaret Chase Smith Center for Public Policy.

Walter T. Schaffer is Research Training Officer and Acting Director of the Office of Extramural Programs within the Office of Extramural Research at the National Institutes of Health, Bethesda, Maryland.

Patrick Simpkins is a systems engineer and human resources manager in the Information Technology Systems Integration unit of NASA's Management Systems Division, Washington, D.C.

Michael S. Teitelbaum is a demographer and program director at the Alfred P. Sloan Foundation, New York, New York.

Katherine K. Wallman is Chief Statistician at the United States Office of Management and Budget, Washington, D.C.

RAND Participants

David M. Adamson is a senior research communicator with the RAND Corporation in Arlington, Virginia.
Gabrielle Bloom is a research assistant with RAND in Arlington, Virginia.
William P. Butz is President and Chief Executive Officer of the Population Reference Bureau in Washington, D.C., and a consultant to RAND.
Stephen J. Carroll is a senior economist with RAND in Santa Monica, California, as well as Secretary and Chief Financial Officer of the Council for Aid to Education.
Terrence K. Kelly is a senior researcher with RAND in Pittsburgh, Pennsylvania.